AMERICAN
JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

REINHOLD BAER
UNIVERSITY OF ILLINOIS

WEI-LIANG CHOW
THE JOHNS HOPKINS UNIVERSITY

SAMUEL EILENBERG
COLUMBIA UNIVERSITY

LAUREL WINTNER
THE JOHNS HOPKINS UNIVERSITY

WITH THE COÖPERATION OF

S. S. CHERN
C. CHEVALLEY
J. A. DIEUDONNÉ
A. M. GLEASON
P. HARTMAN
G. P. HOCHSCHILD
I. KAPLANSKY
E. R. KOLCHIN
W. S. MASSEY
D. C. SPENCER
A. D. WALLACE
ANDRÉ WEIL

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

VOLUME LXXVI

1954

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

OHIO UNIVERSITY
LIBRARY

CLOSED MANIFOLDS WITH HOMOGENEOUS COMPLEX STRUCTURE.*

By HSIEN-CHUNG WANG.

1. Introduction. Nearly all the information about closed complex manifolds comes either from the theory of kählerian manifolds or from that of almost complex manifolds. On one hand, we do not know any topological property of complex manifolds which is not a property of almost complex manifolds; on the other hand, very few examples—those given by Calabi-Eckmann [6]—of simply-connected, closed non-kählerian complex manifolds are known. On account of this situation, it seems desirable firstly to discuss some particular classes of complex manifolds which may impart some ideas for the general case. If we consider the complex structure as a geometrical object, a very natural assumption will be homogeneity. In fact, a complex manifold W is called *homogeneous* if, given any two points of W, there exists an analytic homeomorphism of W carrying one to the other. It is the aim of the present paper to study the simply-connected, closed homogeneous complex manifolds which, for simiplicity, will be called a C-space. We find that there are many more C-spaces than we expect, and the underlying space of most of them has infinitely many inequivalent non-kählerian complex structures.

Concerning the general properties, we have

THEOREM A. *If a C-space has vanishing second Betti number, then its Euler characteristic must vanish.*

THEOREM B. *Let M be a simply-connected closed manifold. If the Euler characteristic χ of M differs from zero, then M has at most a finite number of inequivalent homogeneous complex structures. If $\chi = 0$, then M has either no homogeneous complex structure or non-countably many inequivalent ones.*

THEOREM C. *Let K be a compact semi-simple Lie group, and X a closed connected subgroup whose semi-simple part coincides with the semi-simple part of the centraliser of a toral subgroup of K. Then the coset*

* Received March 18, 1953.

OHIO UNIVERSITY
LIBRARY

space K/X, if even-dimensional, has a homogeneous complex structure. Conversely, each C-space is homeomorphic with such a coset space.

Using S^6 as an example, we see that Theorem A does not hold for homogeneous almost complex manifolds. It is unknown to the author whether or not Theorem A is valid if we drop the homogeneity restriction. Theorem C tells us that the class of C-spaces is rather large.

For the actual determination of the C-spaces, we introduce a class of manifolds, called M-manifolds. In fact, a coset space K/U of a compact simple Lie group K is called an M-manifold if U is the semi-simple part of the centraliser of a toral subgroup of K. All M-manifolds are eventually determined. In this class, we have, in particular, the space of compact simple Lie groups, the complex Stiefel manifolds, the quarternion Stiefel manifolds and the ordinary Stiefel manifolds of the form $V_{m,2k}$. The M-manifolds and C-spaces are related as follows:

THEOREM D. *Each even-dimensional M-manifold has infinitely many inequivalent homogeneous complex structures and so does the product of two odd-dimensional ones. They have vanishing second Betti number, and hence they are non-kählerian. Moreover, each C-space W is homeomorphic with a fibre decomposition space of a product of M-manifolds with torus as the fibre* (i. e., *the product is a torus bundle over the underlying space of W*).

It is to be noted that all these theorems still hold if, instead of simply-connectedness, we asume the spaces to have finite fundamenfal groups. The proof is essentially the same.

2. Complex coset spaces. Let G be a complex Lie group. A subgroup L of G is called a *complex subgroup* if L is closed and the underlying point set of L is a complex submanifold of G (by a complex submanifold J of a complex manifold M, we mean a topological submanifold with the property that, for each point q of M, there exist analytic functions $f_1, f_2, \cdots, f_s$ defined in some neighborhood U of q such that

$$J \cap U = \{p: f_i(p) = 0, i = 1, 2, \cdots, s\}).$$

The identity component of a complex subgroup of a complex Lie group G is generated by a subalgebra (over complex field) of the complex Lie algebra associated with G. Conversely, the subgroup of G generated by such a subalgebra, if closed, is a complex subgroup.

Let L be a complex subgroup of a complex Lie group G. A complex

structure can be defined over the space G/L of left cosets in the usual manner. In fact, let $g \in G$, $q = \pi(g)$ where $\pi: G \to G/L$ denotes the projection, and let $\mathfrak{A}_G(g)$ be the class of analytic local functions on G at g. We set $\mathfrak{A}_{G/L}(q)$ to be the class of all complex-valued functions ϕ defined in some neighborhood of q such that $\phi\pi \in \mathfrak{A}_G(g)$. Using the same arguments as in the real analytic case, we can verify that the classes $\mathfrak{A}_{G/L}(q)$, $q \in G/L$, of local functions defines a complex analytic structure over G/L. The resulting complex manifold will be called the *complex coset space* G/L. With respect to this complex structure, the mapping $g: G/L \to G/L$ given by $g(xL) = gxL$ is complex analytic where g denotes an element of G. Thus the complex coset space G/L is a homogeneous complex manifold.

In general, a homogeneous complex manifold cannot be written as a complex coset space. Nevertheless, if compactness is assumed, we have the following:

(2.1) *Each closed homogeneous complex manifold is analytically homeomorphic with a complex coset space.*

Proof. Let W be a closed homogeneous complex manifold, and B the identity component of the group of all analytic homeomorphisms of W. Bochner and Montgomery [3] have shown that a complex structure can be given to the space of B such that (i) B becomes a complex Lie group with respect to this structure, and (ii) the mapping $T: B \times W \to W$ given by $T(g, x) = g(x)$, $g \in B$, $x \in W$, is analytic. From the homogeneity of W, we see that B is transitive over W. Let L be the isotropic subgroup of G at a point q of W. L is evidently closed. Moreover, from (ii), L is, locally, defined by analytic equations in B. Thus L is a complex subgroup of B, and thus we can form the complex coset space B/L. Since W is a manifold, the mapping $\theta: B/L \to W$ given by $\theta(gL) = g(q)$, $g \in B$, is an onto homeomorphism. In fact, θ is two-sided analytic. To see this, let $T_q: B \to W$ be defined by $T_q(g) = g(q) = T(g, q)$, and let ψ be an analytic local function on W. It is easy to see that $\psi T_q = \psi\theta\pi$ where $\pi: B \to B/L$ denotes the projection. Because of (ii), T_q is an analytic mapping so that ψT_q or $\psi\theta\pi$ is an analytic local function on B. From the definition of complex coset spaces, it follows that $\psi\theta$ is an analytic local function on B/L. This tells us that the mapping θ is analytic. The analicity of its inverse θ^{-1} follows as a consequence [1], p. 179. (2.1) is thus proved.

If the fundamental group of W is assumed to be finite, stronger assertion than (2.1) can be obtained. For this purpose, we shall first show a trivial property of transformation groups.

(2.2) *Let $W = K/L$ be a coset space of a compact Lie group K. If the fundamental group $\pi_1(W)$ is finite, then the maximal semi-simple subgroup J of K is transitive on W.*

Proof. Since K is compact, its maximal semi-simple subgroup J is a normal subgroup, and moreover, K/J is abelian. It follows that JL is a closed normal subgroup and that K/JL is a toral group. Since all the groups involved are Lie groups, we know [11], p. 30, that K/L is a fibre bundle over K/JL with JL/J as the fibre. JL/L is connected as J is connected. Thus the fundamental group $\pi_1(K/JL)$ of the base space is a factor group of the fundamental group $\pi_1(K/L)$ of the bundle space. By hypothesis, $\pi_1(K/L)$ is finite, and thus $\pi_1(K/JL)$ is also finite. On the other hand, we know that K/JL is a toral group. Therefore, K/JL can only contain the identity, or in other words, $K = JL$. This tells us that J is transitive on $W = K/L$.

(2.3) *Let $W = B/L$ be a compact complex coset space. If the fundamental group of W is finite, then each maximal connected semi-simple subgroup G of B is transitive over W. Moreover, W is analytically homeomorphic with the complex coset space $G/(G \cap L)$.*

Proof. It is well-known that such a G is generated by the semi-simple part of a Levi-decomposition of the complex Lie algebra of B. Thus G is a complex subgroup. Malcev [8] has shown that all the G's are conjugate in B. Therefore, the choice of G is immaterial. Let K be a maximal compact subgroup of B. We know that K is transitive over W [9]. It follows from (2.2) that the maximal connected semi-simple subgroup J of K is transitive on W. This J is contained in a maximal connected semi-simple subgroup G of B. Thus G is transitive over W. This proves the first part of (2.3).

To show the second part, let us put $L' = G \cap L$. Since both G and L are complex subgroups, L' is also a complex subgroup, and thus the complex coset space G/L' is well-defined. Because of the transitiveness of G on B/L, the mapping $f: G/L' \to B/L$ defined by $f(gL') = gL$ is an onto homeomorphism. We can easily verify that f is analytic. (2.3) is proved.

3. Solvable subgroups. From the above discussions, the study of compact homogeneous compact spaces is reduced to the study of complex coset spaces G/L. In this section, we shall establish some necessary conditions on the maximal solvable subgroups of L in order that G/L is compact.

It is desirable to say here a few words about our notations. Let G be

a Lie group. $K(G)$ always denotes a maximal connected compact subgroup of G. By dimension of G, we shall always mean its topological dimension. Thus for a complex Lie group G, dim. G is twice of the dimension of its complex Lie algebra.

(3. 1) *Let $W = G/L$ be a coset space of a Lie group G. If*

(3. 2) $$\text{dim. } G - \text{dim. } K(G) = \text{dim. } L - \text{dim. } K(L),$$

then W is compact. Conversely, if W is compact and has a finite fundamental group, then equality (3. 2) *holds.*

Proof. Suppose that (3. 2) holds. Then dim. W = dim. $K(G)$ — dim. $K(L)$. Let R be an orbit in W under $K(G)$. Then R is homeomorphic with $K(G)/(K(G) \cap L)$, and then dim. R = dim. $K(G)$ — dim. $(K(G) \cap L)$. Since $K(G) \cap L$ is a compact subgroup of L, its dimension cannot be greater than that $K(L)$. It follows then

$$\text{dim. } R \geq \text{dim. } K(G) - \text{dim. } K(L) = \text{dim. } W,$$

and then dim. R = dim. W. But R is a closed manifold so that $R = W$. This tells us that W is compact.

Conversely, suppose that W is compact and has finite fundamental group. Choose a $K(G)$ such that $K(G)$ contains a $K(L)$. From a theorem of Montgomery [9], $K(G)$ is transitive on W. Thus

$$\text{dim. } G - \text{dim. } L = \text{dim. } W = \text{dim. } K(G) - \text{dim. } (K(G) \cap L).$$

From our choice of $K(G)$, we known that $K(L)$ must be the identity component of $K(G) \cap L$. Thus dim. $K(L)$ = dim. $(K(G) \cap L)$. This equality together with the preceding one gives (3. 2). (3. 1) is proved.

Now we shall state some known results in a form suitable for our purposes. Let $\boldsymbol{G}$ be a complex semi-simple Lie algebra, and $\boldsymbol{H}$ a Cartan subalgebra of $\boldsymbol{G}$. An element h of $\boldsymbol{H}$ is called *rational* if all the roots of h (i. e., the characteristic roots of $\alpha(h)$ where α denotes the adjoint representation of $\boldsymbol{G}$) are rational numbers. Evidently, h is rational if and only if there exists an integer m such that $\exp m\pi i\alpha(h)$ is the identity. We shall call a base $h_1, h_2, \cdots, h_l$ of $\boldsymbol{H}$ a *rational base* if each h_j is rational. The existence of rational base is well-known. Let $\lambda_1, \lambda_2, \cdots, \lambda_l$ be the coordinates in $\boldsymbol{H}$ with respect to a rational base $h_1, h_2, \cdots, h_l$. Let ϕ be a linear representation of the Lie algebra $\boldsymbol{G}$. Then each weight Λ of ϕ can be written as a linear combination of the λ's: $\Lambda = r_1\lambda_1 + r_2\lambda_2 + \cdots + r_l\lambda_l$. Since the

base is rational, the coefficients r_j must be rational numbers. Thus we can order the weights in the classical manner. In fact, let

$$\Lambda = r_1\lambda_1 + \cdots + r_l\lambda_l, \qquad \Lambda' = s_1\lambda_1 + \cdots + s_l\lambda_l$$

be two weights. We say that $\Lambda < \Lambda'$ if, for some i, we have $r_1 = s_1$, $r_2 = s_2, \cdots, r_i = s_i$, $r_{i+1} < s_{i+1}$. If, in particular, ϕ is the adjoint representation of $\boldsymbol{G}$. Then the weights become roots of $\boldsymbol{G}$. Thus the roots have an ordering. Let e_ρ, e_μ be two root vectors corresponding to the roots ρ, μ respectively. Then we say $e_\rho < e_\mu$ if and only if $\rho < \mu$. It should be noted that this ordering depends on the choice of the rational base of $\boldsymbol{H}$.

Let $\boldsymbol{N}$ be the subspace spanned by all the root vectors $e_{\rho_1}, e_{\rho_2}, \cdots, e_{\rho_t}$ which are greater than zero, and $\boldsymbol{S} = \boldsymbol{N} + \boldsymbol{H}$. A direct verification tells us that $\boldsymbol{S}$ is a maximal solvable subalgebra of $\boldsymbol{G}$. Since $0 < e_\rho$ if and only if $e_{-\rho} < 0$, we know that the number t of positive root vectors is equal to one half of the total number of root vectors. Thus

$$\dim. \boldsymbol{S} = t + l = (\dim. \boldsymbol{G} + l)/2$$

where l denotes the rank of $\boldsymbol{G}$. Morosoff [10] has shown that each solvable subalgebra of $\boldsymbol{G}$ is conjugate to a subalgebra of $\boldsymbol{S}$. Thus all maximal solvable subalgebra of $\boldsymbol{G}$ are conjugate. Moreover, each maximal solvable subalgebra of $\boldsymbol{G}$ can be constructed in the above manner by choosing a suitable $\boldsymbol{H}$ and a suitable rational base of $\boldsymbol{H}$.

Now let G a complex Lie group corresponding to the Lie algebra $\boldsymbol{G}$. We consider the subgroup S generated by $\boldsymbol{S}$ in G. Since $\boldsymbol{S}$ contains a Cartan subalgebra $\boldsymbol{H}$ of $\boldsymbol{G}$, S must be closed [7], and hence a complex subgroup of G. Obviously, S is a maximal connected solvable subgroup, and

$$(3.3) \qquad \dim. S = 2 \dim. \boldsymbol{S} = \dim. \boldsymbol{G} + l = l + (\dim. G)/2.$$

Since S is solvable, each $K(S)$ is a toral group. From the fact that all maximal connected solvable subgroups of G are conjugate to S, it follows that $K(S)$ is a maximal toral subgroup of G. Thus $\dim. K(S) = l$. This equality together with (3.3) gives

$$(3.4) \quad \dim. G - \dim. K(G) = (\dim. G)/2 = \dim. S - \dim. K(S).$$

Now we can show easily the following:

(3.5) *Let B be a complex Lie group. All its maximal connected solvable subgroups M are complex subgroups and are conjugate to one another. Moreover, $\pi_1(B/M)$ is finite and*

$$\dim. B - \dim. K(B) = \dim. M - \dim. K(M).$$

Proof. Let $B = GR$ be a Levi-decomposition of B where G denotes a maximal semi-simple subgroup of B and R the radical. Evidently, each maximal connected solvable subgroup M of B takes the form $M = SR$ where S is a maximal connected solvable subgroup of G. Thus M is a complex subgroup of B. From the results of Morosoff, we know that all such S's are conjugate in G. It follows that all such M's are conjugate in B. The first part of (3. 5) is proved. For the second part, we consider the coset space $W = B/M$. Since R is a common normal subgroup of B and M, we have $W = (B/R)/(M/R) = G/S$. (3. 4) together with (3. 1) tells us that $W = B/M$ is compact. G is a complex semi-simple Lie group so that its fundamental group is finite. From this fact and the connectedness of S, it follows that the fundamental group of W is finite. Now applying the second part of (3. 1) to W, we get the equality

$$\dim. B - \dim. K(B) = \dim. M - \dim. K(M).$$

This proves (3. 5).

(3. 6) *Let $W = G/L$ be a complex coset space of a complex semi-simple Lie group G, and M a maximal connected solvable subgroup of L. If W is compact and has finite fundamental group, then* $\dim. M = (\dim. G)/2 + v$ *where v denotes the dimension of a maximal compact subgroup of M.*

Proof. From (3. 1), we know that

$$\dim. L - \dim. K(L) = \dim. G - \dim. K(G).$$

Proposition (3. 5) tells us that

$$\dim. L - \dim. K(L) = \dim. M - \dim. K(M).$$

Thus $\dim. M - \dim. K(M) = \dim. G - \dim. K(G)$. Here G is a complex semi-simple Lie group so that $\dim. K(G) = (\dim. G)/2$. It follows then $\dim. M = (\dim. G)/2 + \dim. K(M)$. (3. 6) is proved.

4. *J*-subalgebras. In this section, we shall introduce and study a class of subalgebras which facilitate to determine the structure of the isotropic subgroup L of the compact complex coset spaces G/L. Let $\boldsymbol{H}$ be a Cartan subalgebra of a complex semi-simple Lie algebra $\boldsymbol{G}$. A linear subspace $\boldsymbol{P}$ of dimension p in $\boldsymbol{H}$ is called a *rational plane* or a *rational subalgebra* if $\boldsymbol{P}$ contains p linearly independent rational elements. If $p = l - 1$ where $l = \text{rank } (\boldsymbol{G}) = \dim. \boldsymbol{H}$, then we call it a *rational hyperplane.* Now let $\boldsymbol{P}$ be a p-dimensional rational plane. We choose p linearly independent rational

elements $k_1, \cdots, k_p$ in $\boldsymbol{P}$. Let α denote the adjoint representation of $\boldsymbol{G}$. With respect to a suitable base of $\boldsymbol{G}$, each $\alpha(k_j)$ takes the form of a diagonal matrix. From the definition of rational elements in $\boldsymbol{H}$, we know that the elements of these diagonal matrices $\alpha(k_j)$ must be rational numbers. It follows that, in the adjoint group G^* of $\boldsymbol{G}$, $\boldsymbol{P}$ generates a closed subgroup isomorphic with the direct product $T^p \times R^p$ of a p-dimensional toral group and a p-dimensional vector group. Since each group form of the complex semi-simple Lie algebra $\boldsymbol{G}$ is a covering group of G^* with *finite* leaves, the above property of $\boldsymbol{P}$ not only holds for G^* but also for any group form of $\boldsymbol{G}$.

DEF. *A subalgebra $\boldsymbol{J}$ of $\boldsymbol{H}$ is called a J-subalgebra if it does not lie on any rational hyperplane of $\boldsymbol{H}$.*

(4.1) *Let $\boldsymbol{H}$ be a Cartan subalgebra of a complex semi-simple Lie algebra $\boldsymbol{G}$ of rank l. Let $\boldsymbol{J}$ be a subalgebra of $\boldsymbol{H}$, and J the subgroup generated by $\boldsymbol{J}$ in a group form G of $\boldsymbol{G}$. If J contains an l-dimensional vector subgroup, then $\boldsymbol{J}$ is a J-subalgebra.*

Proof. If $\boldsymbol{J}$ lies on a rational hyperplane $\boldsymbol{P}$, then J would be a subgroup of the group P generated by $\boldsymbol{P}$ in G. We know that P is isomorphic with the direct product $T^{l-1} \times R^{l-1}$ of an $(l-1)$-dimensional group and an $(l-1)$-dimensional vector group. P has no l-dimensional vector subgroup. A contradiction is obtained. Hence $\boldsymbol{J}$ is a J-subalgebra.

Before establishing further properties of J-subalgebras, we have to digress a little and define some terminiologies used in representation theory. Let ϕ be a representation of the Lie algebra $\boldsymbol{G}$, and V the complex vector space on which $\phi(\boldsymbol{G})$ takes place. We shall fix a Cartan subalgebra $\boldsymbol{H}$ of $\boldsymbol{G}$ and a rational base $h_1, \cdots, h_l$ of $\boldsymbol{H}$. The coordinates with respect to this base of $\boldsymbol{H}$ will be denoted by $\lambda_1, \cdots, \lambda_l$. Associated with ϕ, there are a finite number of linear forms $\Lambda_1, \Lambda_2, \cdots, \Lambda_s$ over $\boldsymbol{H}$, called weights. To each weight Λ corresponds a linear subspace V_Λ, called weight space, such that, for each vector η of V_Λ and each element h of $\boldsymbol{H}$, we have $\phi(h)\eta = \Lambda(h)\eta$. Moreover, the space V is decomposed into the direct sum $V = V_{\Lambda_1} + V_{\Lambda_2} + \cdots + V_{\Lambda_s}$ of the weight spaces. Since we have a fixed coordinate system $\lambda_1, \cdots, \lambda_l$ in $\boldsymbol{H}$, we can write the weights as linear combinations of the λ's:

$$\Lambda_j = \sum_{k=1}^{l} m_{jk}\lambda_k, \qquad j = 1, 2, \cdots, s.$$

The base $h_1, \cdots, h_l$ of $\boldsymbol{H}$ is rational so that the coefficients m_{jk} are all rational numbers.

(4.2) *Let* $\boldsymbol{J} \subset \boldsymbol{H}$ *be a J-subalgebra and* V^* *be linear subspace of* V *invariant under* $\phi(\boldsymbol{J})$. *Then*

$$V^* = (V^* \cap V_{\Lambda_1}) + (V^* \cap V_{\Lambda_2}) + \cdots + (V^* \cap V_{\Lambda_s}).$$

Here all the notations have the same meaning as in the preceding paragraph.

Proof. Let Λ_i and Λ_j be two distinct weights. We shall first see that $\Lambda_i(h) \not\equiv \Lambda_j(h)$, $h \in \boldsymbol{J}$. If they were identical for all h in $\boldsymbol{J}$, then $\boldsymbol{J}$ would lie in the hyperplane $\boldsymbol{P}$ which is defined in terms of the coordinates $\lambda_1, \cdots, \lambda_l$ by

$$\sum_{k=1}^{l} (m_{ik} - m_{jk})\lambda_k = 0.$$

Since $h_1, \cdots, h_l$ form a rational base of $\boldsymbol{H}$ and the m's are rational numbers, this $\boldsymbol{P}$ is a rational hyperplane. This contradicts our assumption that $\boldsymbol{J}$ is a J-subalgebra. Thus $\Lambda_i(h) \not\equiv \Lambda_j(h)$, $h \in \boldsymbol{J}$. We have only a finite number of weights so that there exists an element h_0 of $\boldsymbol{J}$ such that $\Lambda_i(h_0) \neq \Lambda_j(h_0)$ whenever $i \neq j$.

Now let ξ be a vector of V^*. It can be written in the form:

$$\xi = \xi_1 + \xi_2 + \cdots + \xi_s, \quad \xi_j \in V_{\Lambda_j}.$$

Thus

$$(\phi(h_0))^k \xi = \sum_{j=1}^{s} (\Lambda_j(h_0))^k \xi_j, \qquad k = 0, 1, 2, \cdots.$$

Since $\Lambda_i(h_0) \neq \Lambda_j(h_0)$ for $i \neq j$, the vectors $\xi_1, \xi_2, \cdots, \xi_s$ can be written as linear combinations of

$$\xi, \phi(h_0)\xi, (\phi(h_0))^2\xi, \cdots, (\phi(h_0))^{s-1}\xi.$$

But on account of the invariance of V^* under $\phi(\boldsymbol{J})$, the latter vectors belong to V^*. It follows then $\xi_j \in V^*$, and then $\xi_j \in V^* \cap V_{\Lambda_j}$. Proposition (4.2) is therefore proved.

(4.3) *Let* $\boldsymbol{L}$ *be a subalgebra of* $\boldsymbol{G}$ *such that* $\boldsymbol{L}$ *contains a J-subalgebra* $\boldsymbol{J}$ *of* $\boldsymbol{H}$. *Then* $\boldsymbol{L}$ *is spanned by* $\boldsymbol{L} \cap \boldsymbol{H}$ *and some root vectors of* $\boldsymbol{G}$.

Proof. In (4.2), let us take ϕ to be the adjoint representation of $\boldsymbol{G}$. Then $V = \boldsymbol{G}$, and the weights of ϕ are 0 and the various roots of $\boldsymbol{G}$. The weight space V_0 corresponding to zero is $\boldsymbol{H}$, and the weight space V_ρ corresponding to a root ρ of $\boldsymbol{G}$ is the one-dimensional space containing the root vector e_ρ. Taking $V^* = \boldsymbol{L}$, we get (4.3) immediately from (4.2).

5. The structure of the isotropic subgroup. Let $W = G/L$ be a complex coset space of a complex semi-simple Lie group G. We shall always assume that W is compact and has finite fundamental group. Let M be a maximal connected solvable subgroup of L. Choose a maximal connected solvable subgroup S of G such that $M \subset S$. Let $\boldsymbol{G}$, $\boldsymbol{L}$, $\boldsymbol{S}$, $\boldsymbol{M}$ be the complex Lie algebras of G, L, S, M respectively. From the facts that $\boldsymbol{M} \subset \boldsymbol{S}$ and that $\boldsymbol{M}$ is a maximal solvable subalgebra of $\boldsymbol{L}$, we see immediately that $\boldsymbol{M} = \boldsymbol{L} \cap \boldsymbol{S}$. From § 3, there exists a Cartan subalgebra $\boldsymbol{H}$ of $\boldsymbol{G}$ and a rational base $h_1, \cdots, h_l$ of $\boldsymbol{H}$ such that $\boldsymbol{S}$ is spanned by $\boldsymbol{H}$ and all the positive root vectors $e_{\rho_1}, e_{\rho_2}, \cdots, e_{\rho_t}$ of $\boldsymbol{G}$ (the ordering of the root vectors is defined with respect to the base $h_1, \cdots, h_l$). Here $l = \text{rank}(G)$ and dim. $G = 2t + l$. From (3. 6), we know

$$\text{dim. } M = (\text{dim. } G)/2 + v = \text{dim. } \boldsymbol{G} + v = 2t + l + v, \quad v = \text{dim. } K(M).$$

Thus

$$\text{dim. } \boldsymbol{M} = t + (l + v)/2. \tag{5. 1}$$

Let $\boldsymbol{J} = \boldsymbol{H} \cap \boldsymbol{M}$. Since both $\boldsymbol{M}$ and $\boldsymbol{H}$ are linear subspaces of $\boldsymbol{S}$,

$$\text{dim. } \boldsymbol{J} \geq \text{dim. } \boldsymbol{M} + \text{dim. } \boldsymbol{H} - \text{dim. } \boldsymbol{S} = (l + v)/2.$$

Let J and H be the subgroups generated by $\boldsymbol{J}$ and $\boldsymbol{H}$ in G respectively. H is evidently closed. The subgroup J, being the identity component of the intersection $M \cap H$ of two closed subgroups, is also closed. Moreover,

$$\text{dim. } J = 2 \text{ dim. } \boldsymbol{J} \geq l + v.$$

We can always write the connected abelian Lie group J in the product form: $J = T^a \times R^b$ where T^a denotes an n-dimensional toral group and R^b a b-dimensional vector group. Evidently, $a + b = \text{dim. } J \geq l + v$. Since $T^a \subset J \subset M$, $a \leq \text{dim. } K(M) = v$. It follows then $b \geq l$. On the other hand, the subgroup H, being generated by a Cartan subalgebra, is isomorphic with the direct product $T^l \times R^l$ of an l-dimensional toral group and an l-dimensional vector group. Hence H, and hence J, cannot have any vector subgroup of dimension higher than l. Thus $b = l$, $a = v$ and $J = T^v \times R^l$. From (4. 1), we know that $\boldsymbol{J}$ must be a J-subalgebra. (4. 3) then tells us that $\boldsymbol{L}$ is spanned by $\boldsymbol{L} \cap \boldsymbol{H}$ and some root vectors of $\boldsymbol{G}$. Now we shall see that $\boldsymbol{L}$ contains all the positive root vectors. For this purpose, let us consider $\boldsymbol{M}$. Since $\boldsymbol{J} \subset \boldsymbol{M}$, $\boldsymbol{M}$ is spanned by $\boldsymbol{M} \cap \boldsymbol{H}$ and some root vectors, say $e_{\alpha_1}, \cdots, e_{\alpha_s}$. Then dim. $\boldsymbol{M} = \text{dim.}(\boldsymbol{M} \cap \boldsymbol{H}) + s = \text{dim. } \boldsymbol{J} + s = s + (l + v)/2$. Comparing this equality with (5. 1), we get $s = t$. Since $\boldsymbol{M} \subset \boldsymbol{S}$, all these

root vectors $e_{\alpha_1}, \cdots, e_{\alpha_s}$ must be positive. But we have only t positive root vectors. Thus $\boldsymbol{M}$, and hence $\boldsymbol{L}$, contains all the positive root vectors.

Summing up the above results, we have

(5.2) *Let $W = G/L$ be a complex coset space of a complex semi-simple Lie group G. Suppose that W is compact and has finite fundamental group. Let $\boldsymbol{G}$, $\boldsymbol{L}$ be the complex Lie algebras of G, L respectively. Then there exists a Cartan subalgebra $\boldsymbol{H}$ of $\boldsymbol{G}$ and a rational base $h_1, \cdots, h_l$ of $\boldsymbol{H}$ with the following properties*: (i) *$\boldsymbol{L} \cap \boldsymbol{H}$ is a J-subalgebra of $\boldsymbol{H}$,* (ii) *$\boldsymbol{L}$ is spanned by $\boldsymbol{L} \cap \boldsymbol{H}$ and some root vectors of $\boldsymbol{G}$, and* (iii) *all the root vectors which are positive with respect to the base $h_1, \cdots, h_l$ lie on $\boldsymbol{L}$.*

To study further properties of $\boldsymbol{L}$, let us denote by Θ the totality of roots β of $\boldsymbol{G}$ such that both e_β and $e_{-\beta}$ belong to $\boldsymbol{L}$. The roots in Θ occur in pairs $\beta, -\beta$.

(5.3) *The linear subspace $\boldsymbol{Q}$ spanned by*

$$\{e_\beta, h_\beta = [e_\beta, e_{-\beta}] : \beta \in \Theta\}$$

is a semi-simple subalgebra of $\boldsymbol{L}$.

Proof. Since $[h_\beta, \boldsymbol{Q}] \subset \boldsymbol{Q}$, to show that $\boldsymbol{Q}$ is a subalgebra, it suffices to show that $[e_\beta, e_\gamma] \in \boldsymbol{Q}$ for $\beta, \gamma \in \Theta$. If $\beta + \gamma$ is not a root of $\boldsymbol{G}$, then this evidently holds. Now assume $\beta + \gamma = \tau$ to be actually a root of $\boldsymbol{G}$. Since e_β and e_γ belong to $\boldsymbol{L}$, the root vector e_τ belongs to $\boldsymbol{L}$ as well. By definition of Θ, we know that $e_{-\beta}, e_{-\gamma} \in \boldsymbol{L}$. It follows then $[e_{-\beta}, e_{-\gamma}] \in \boldsymbol{L}$, and then $e_{-\tau} \in \boldsymbol{L}$. This tells us that $\tau \in \Theta$, or in other words, $e_\tau \in \boldsymbol{Q}$. But e_τ and $[e_\beta, e_\gamma]$ only differ by a non-zero scalar. Thus $[e_\beta, e_\gamma] \in \boldsymbol{Q}$. Thus $\boldsymbol{Q}$ forms a subalgebra. Its semi-simplicity is evident.

In fact, $\boldsymbol{Q}$ is the essential part of $\boldsymbol{L}$. The following property of $\boldsymbol{Q}$ plays a basic rôle in our study of compact homogeneous complex manifolds.

(5.4) *There exists in $\boldsymbol{H}$ a rational subalgebra $\boldsymbol{H}_1$ such that the centraliser $\boldsymbol{C}$ of $\boldsymbol{H}_1$ is decomposed into the direct sum $\boldsymbol{H}_1 \oplus \boldsymbol{Q}$ of $\boldsymbol{H}_1$ and $\boldsymbol{Q}$.*

Proof. Let $\boldsymbol{H}_1$ be the totality of elements h of $\boldsymbol{H}$ such that $[h, \boldsymbol{Q}] = 0$. It can be easily verified that

$$\boldsymbol{H}_1 = \{h : h \in \boldsymbol{H}, \beta(h) = 0 \text{ for all } \beta \text{ in } \Theta\}.$$

Evidently $\boldsymbol{H}_1$ is a subalgebra of $\boldsymbol{H}$. To see that it is rational, let us put, for each $\beta \in \Theta$,

$$\boldsymbol{P}_\beta = \{h : h \in \boldsymbol{H}, \beta(h) = 0\}.$$

This $\boldsymbol{P}_\beta$ is a rational hyperplane. Since $\boldsymbol{H}_1$ is the intersection of all the $\boldsymbol{P}_\beta$, $\beta \in \boldsymbol{H}$, it follows that $\boldsymbol{H}_1$ is rational.

Let l^* be the rank of $\boldsymbol{Q}$. Then among the roots of $\boldsymbol{Q}$, or what is the same, among the roots β in Θ, there are exactly l^* independent ones. Thus among the rational hyperplanes $\boldsymbol{P}_\beta$, there are l^* independent ones, and thus dim. $\boldsymbol{H}_1 =$ dim. $(\cap \boldsymbol{P}_\beta) = l - l^*$. Now let $\boldsymbol{H}_2$ be the space spanned by $\{h_\beta = [e_\beta, e_{-\beta}] : \beta \in \Theta\}$. This is evidently a Cartan subalgebra of $\boldsymbol{Q}$, and thus dim. $\boldsymbol{H}_2 = l^*$. Moreover, since $\boldsymbol{Q}$ is semi-simple and $[\boldsymbol{H}_1, \boldsymbol{Q}] = 0$, we have

$$\boldsymbol{H}_1 \cap \boldsymbol{H}_2 = \boldsymbol{H}_1 \cap \boldsymbol{Q} = 0.$$

A counting of dimensions of $\boldsymbol{H}$, $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ tells us at once that $\boldsymbol{H} = \boldsymbol{H}_1 + \boldsymbol{H}_2$.

Let $\boldsymbol{C}$ be the centraliser of $\boldsymbol{H}_1$ in $\boldsymbol{G}$. Obviously, $\boldsymbol{H} \subset \boldsymbol{C}$ and $\boldsymbol{Q} \subset \boldsymbol{C}$. Thus to prove (5. 4), it is sufficient to prove $\boldsymbol{C} \subset \boldsymbol{H}_1 + \boldsymbol{Q}$. We shall prove this by using method of contradiction. Suppose $\boldsymbol{C} \not\subset \boldsymbol{H}_1 + \boldsymbol{Q}$. Since $\boldsymbol{C}$ contains a Cartan subalgebra $\boldsymbol{H}$, $\boldsymbol{C}$ is spanned by $\boldsymbol{H}$ and certain root vectors of $\boldsymbol{G}$. But we already know that $\boldsymbol{H} = \boldsymbol{H}_1 + \boldsymbol{H}_2 \subset \boldsymbol{H}_1 + \boldsymbol{Q}$. Thus there must be a root vector e_ρ of $\boldsymbol{G}$ such that $e_\rho \in \boldsymbol{C}$, $e_\rho \notin \boldsymbol{Q}$. From the definition of $\boldsymbol{C}$ and $\boldsymbol{Q}$, we see at once that $e_{-\rho}$ also has this property, i. e., $e_{-\rho} \in \boldsymbol{C}$, $e_{-\rho} \notin \boldsymbol{Q}$. Thus there exists always a *positive* root vector of $\boldsymbol{G}$ which belongs to $\boldsymbol{C}$ but not to $\boldsymbol{Q}$. Let $\boldsymbol{D}$ be the linear subspace spanned by all such positive root vectors. Then $\boldsymbol{D} \neq 0$, $\boldsymbol{D} \subset \boldsymbol{C}$, $\boldsymbol{D} \cap \boldsymbol{Q} = 0$. Moreover, $\boldsymbol{D} \subset \boldsymbol{L}$ on account of (5. 2). Firstly, we want to show $[\boldsymbol{Q}, \boldsymbol{D}] \subset \boldsymbol{D}$. Since $[\boldsymbol{H}_2 \boldsymbol{D}] \subset \boldsymbol{D}$, it suffices to show that, for each $e_\beta \in \boldsymbol{Q}$ and each $e_\rho \in \boldsymbol{D}$, $[e_\beta, e_\rho] \in \boldsymbol{D}$. If $\beta + \rho$ is not a root of $\boldsymbol{G}$, then $[e_\beta, e_\rho] = 0$ evidently belongs to $\boldsymbol{D}$. Now suppose $\beta + \rho = \tau$ to be a root of $\boldsymbol{G}$. Since $e_\beta, e_\rho \in \boldsymbol{L}$, $e_\tau \in \boldsymbol{L}$. The vector $e_{-\tau}$ cannot belong to $\boldsymbol{L}$, for if so, we would have $\tau \in \Theta$, $e_\tau \in \boldsymbol{Q}$, $[e_\tau, e_{-\beta}] \in \boldsymbol{Q}$, and then $e_\rho \in \boldsymbol{Q}$ which is impossible. Thus $e_{-\tau} \notin \boldsymbol{L}$. From this fact, it follows that e_τ is positive and $e_\tau \notin \boldsymbol{Q}$. We know that $e_\beta, e_\rho \in \boldsymbol{C}$ so that e_τ which is a multiple of $[e_\beta, e_\rho]$ must belong to $\boldsymbol{C}$. Thus $e_\tau \in \boldsymbol{D}$, and hence $[e_\beta, e_\rho] \in \boldsymbol{D}$. This proves that $[\boldsymbol{Q}, \boldsymbol{D}] \subset \boldsymbol{D}$.

Let ϕ be the representation of the complex semi-simple Lie algebra $\boldsymbol{Q}$ over the linear space $\boldsymbol{D}$ defined by

$$\phi(x)(y) = [x, y], x \in \boldsymbol{Q}, y \in \boldsymbol{D}.$$

On account of the fact $[\boldsymbol{Q}, \boldsymbol{D}] \subset \boldsymbol{D}$, this representation is well-defined. Let $e_{\rho_1}, e_{\rho_2}, \cdots, e_{\rho_m}$ be the positive root vectors which span $\boldsymbol{D}$. For each h'' of $\boldsymbol{H}_2$, we have

$$\phi(h'')(e_{\rho_i}) = \rho_i(h'') e_{\rho_i}.$$

Since $\boldsymbol{H}_2$ is a Cartan subalgebra of $\boldsymbol{Q}$, e_{ρ_i} are weight vectors of the representation ϕ. Their weights are the linear forms ρ_i confined to $\boldsymbol{H}_2$. It is well-known that the sum of all the weights of a representation of a complex semi-simple Lie algebra is equal to zero. Thus, for each h'' of $\boldsymbol{H}_2$, we have $\sum_{i=1}^{m} \rho_i(h'') = 0$. Let h be a general element of $\boldsymbol{H}$. It can be written in the form $h = h' + h''$ where $h' \in \boldsymbol{H}_1$, $h'' \in \boldsymbol{H}$. Thus

$$\sum_{i=1}^{m} \rho_i(h) = \sum_{i=1}^{m} \rho_i(h') + \sum_{i=1}^{m} \rho_i(h'') = \sum_{i=1}^{m} \rho_i(h').$$

But each e_{ρ_i} belongs to the centraliser $\boldsymbol{C}$ of $\boldsymbol{H}_1$. Thus $\rho_i(h') = 0$, and thus

$$\sum_{i=1}^{m} \rho_i(h) = 0, \qquad h \in \boldsymbol{H}.$$

From the definition of $\boldsymbol{D}$, all the root vectors $e_{\rho_1}, \cdots, e_{\rho_m}$ are positive in the sense of § 3 so that $\sum_{i=1}^{m} \rho_i(h)$ cannot vanish for all h in $\boldsymbol{H}$. A contradiction is therefore obtained. This proves (5. 4).

6. The underlying space of homogeneous complex manifolds. Now we are in position to give some definite results about homogeneous complex manifolds. To simplify the statements, we find it desirable to lay down here two definitions.

DEF. A simply-connected, compact homogeneous complex manifold will be called a *C-space*.

DEF. Let K be a compact semi-simple Lie group. By a *C-subgroup* of K, we mean a closed and connected subgroup whose semi-simple part coincides with the semi-simple part of the centraliser of a toral subgroup of K. Obviously, the identity is a C-subgroup, each toral subgroup is a C-subgroup and the centraliser of a toral subgroup is a C-subgroup.

THEOREM I. *Each C-space is homeomorphic with a coset space K/X where K denotes a compact semi-simple Lie group, and X a C-subgroup of K.*

Proof. Let W be a C-space. From (2. 3), we can regard W as a complex coset space G/L of a complex semi-simple Lie group G. Suppose that $\boldsymbol{G}, \boldsymbol{L}, \boldsymbol{H}, \boldsymbol{H}_1, \boldsymbol{C}, \Theta, h_1, \cdots, h_l$ have the same meaning as in § 5. From now on, both real Lie algebra and complex Lie algebra will occur in our discussions. To distinguish them, we shall always use an 'o' as upper index in each real Lie algebra. It is known that the root vectors of $\boldsymbol{G}$ can be so normalised that

$i[e_\rho, e_{-\rho}]$ are rational elements and $[e_\alpha, e_\beta]$ is a rational multiple of $e_{\alpha+\beta}$ whenever $\alpha+\beta$ is a root [13]. From now on, all the root vectors e_ρ are assumed to be normalised. Let $\boldsymbol{K}^0$ be the Lie algebra consisting of all the linear combinations, with real coefficients, of the following vectors:

$$\{ih_1, \cdots, ih_l, i(e_\rho + e_{-\rho}), e_\rho - e_{-\rho}: \rho = \text{roots of } \boldsymbol{G}\}.$$

The subgroup K generated by this real Lie algebra K^0 is a maximal compact subgroup of G. By Montgomery theorem [9], K is transitive on W. Thus W is homeomorphic with the coset space $K/(K \cap L)$. Let $X = K \cap L$. To prove our theorem, it suffices to show that X is a C-subgroup of K. Firstly, we find that the Lie algebra $\boldsymbol{X}^0$ of X is equal to $\boldsymbol{K}^0 \cap \boldsymbol{L}$, and thus is spanned, in the sense of vector space over real field, by

$$\{\boldsymbol{H}^0 \cap \boldsymbol{L}, i(e_\beta + e_{-\beta}), e_\beta - e_{-\beta}: \beta \in \Theta\} \tag{6.1}$$

where $\boldsymbol{H}^0$ denotes $\boldsymbol{K}^0 \cap \boldsymbol{H}$.

Now consider the subalgebra $\boldsymbol{H}_1$ of $\boldsymbol{H}$ discussed in (5.4). Since it is a rational subalgebra of $\boldsymbol{H}$, it has a rational base $k_1, \cdots, k_a$, $a = \dim. \boldsymbol{H}_1$. Thus the real Lie algebra $\boldsymbol{H}^0_1 = \boldsymbol{H}_1 \cap \boldsymbol{K}^0$ is spanned, over the field of real numbers, by $ik_1, \cdots, ik_a$. It follows then that (a) $\boldsymbol{H}^0_1$ generates a toral subgroup T_1 of K, (b) the centraliser of $\boldsymbol{H}^0_1$ in $\boldsymbol{K}^0$ coincides with the intersection $\boldsymbol{K}^0 \cap \boldsymbol{C}$ of $\boldsymbol{K}^0$ with the centraliser $\boldsymbol{C}$ of $\boldsymbol{H}_1$ in $\boldsymbol{G}$. From § 5, we can easily see that $\boldsymbol{K}^0 \cap \boldsymbol{C}$ is spanned, over the real field, by

$$\{\boldsymbol{H}^0, i(e_\beta + e_{-\beta}), e_\beta - e_{-\beta}: \beta \in \Theta\}.$$

This together with (6.1) tells us that $\boldsymbol{K}^0 \cap \boldsymbol{C}$ and $\boldsymbol{X}^0$ have the same semi-simple part. Thus the centraliser of $\boldsymbol{H}^0_1$ in $\boldsymbol{K}^0$ and $\boldsymbol{X}^0$ have the same semi-simple part. Passing this relation to their respective groups, we obtain the fact that X and the centraliser of T_1 in K have the same semi-simple part. Moreover, because the simply-connectedness of W, X is connected. Thus X is a C-subgroup of K. This proves our theorem.

COROLLARY. *Let W be a C-space containing more than one point. If W has vanishing second Betti number, then its Euler characteristic also vanishes.*

Proof. Since W is simply-connected, the second homotopy group $\pi_2(W)$ of W is isomorphic with the 2-dimensional homology group of W with integer coefficients. Thus $\pi_2(W)$ is finite. From Theorem I, W is homeomorphic with the coset space K/X where K is a compact semi-simple Lie group and X a C-subgroup of K. Thus $\pi_1(K)$ is finite. From the finiteness of both

$\pi_2(W)$ and $\pi_1(K)$, it follows that the fundamental group $\pi_1(X)$ of X is finite. Hence X must be semi-simple. X is a C-subgroup so that X is the semi-simple part of the centraliser U of a certain toral subgroup T_1 of K. Suppose that the Euler characteristic of W does not vanish. Then X and K have the same rank. Such a subgroup X has discrete centraliser in K. Thus T_1 contains only the identity. It follows then that $K = U = X$. In other words, $W = K/X$ consists of a single point. The corollary is therefore proved.

7. Converse of Theorem I. It is the purpose of this section and the next to prove the following theorem which can be regarded as the converse of Theorem I.

THEOREM II. *Let X be a C-subgroup of a simply-connected compact semi-simple Lie group K. If K/X is even-dimensional, then K/X has a homogeneous complex structure, or in other words, K/X is homeomorphic with a C-space.*

We shall first prove a few propositions as lemmas.

(7.1) *Let K be a compact semi-simple Lie group, and X a C-subgroup of K with semi-simple part U. Then there exists a toral subgroup T_1 of K having the following properties:*

(i) *T_1 contains the identity component T_0 of the center of X,*

(ii) *U is the semi-simple part of the centraliser of T_1,*

(iii) $\dim. T_1 + \text{rank } (U) = \text{rank } (K)$.

Proof. Let N be the centraliser of U. Then $T_0 \subset N$, and then T_0 is contained in a maximal toral subgroup, say T_1, of N. We shall see that this T_1 has all the required properties. By hypothesis, X is a C-subgroup so that U is the semi-simple part of the centraliser B of a toral subgroup, say T, of K. Evidently, B and K have the same rank. Let T' be the identity component of the center of B. We have then

$$\dim. T' + \text{rank } (U) = \text{rank } (B) = \text{rank } (K).$$

Since $T' \subset N$, there exists an element p of N such that $pT'p^{-1} \subset T_1$ because T_1 is a maximal toral subgroup of N. It follows then that $\dim. T' \leq \dim. T_1$, and then that $\dim. T_1 + \text{rank } (U) \geq \text{rank } (K)$. On the other hand, T_1U forms a subgroup of K and, because of the semi-simplicity of U, the intersection $T_1 \cap U$ is discrete. Thus

$$\text{rank } (K) \geq \text{rank } (T_1U) = \dim. T_1 + \text{rank } (U).$$

This inequality together with the preceding one tells us that

$$\text{rank } (K) = \dim. T_1 + \text{rank } (U).$$

Therefore, the property (iii) is established.

Since property (i) is evident from our construction, we shall only consider the property (ii). We first observe that $T \subset T'$. Thus $pTp^{-1} \subset T_1$. From the definition of T, it follows that pUp^{-1} is the semi-simple part of the centraliser of pTp^{-1}. On account of the fact $pTp^{-1} \subset T_1$, we know that pUp^{-1} contains the semi-simple part U' of the centraliser of T_1. But $p \in N$ so that $pUp^{-1} = U$. This tells us that $U \supset U'$. On the other hand, we already know that U is semi-simple and that U belongs to the centraliser of T_1. Hence $U \subset U'$, and hence $U = U'$. (7. 1) is hereby proved.

(7. 2) *Let G be a complex semi-simple Lie group, $\boldsymbol{G}$ the complex Lie algebra of G, $\boldsymbol{H}$ a Cartan subalgebra of G. Let T^m be an m-dimensional toral subgroup of G such that its Lie algebra (a Lie algebra over real field) $\boldsymbol{H}^0{}_1$ is contained in $\boldsymbol{H}$. Then there exists m linearly independent rational elements $h_1, h_2, \cdots, h_m$ of $\boldsymbol{H}$ such that $ih_1, ih_2, \cdots, ih_m$ span, with real coefficients, the algebra $\boldsymbol{H}^0{}_1$.*

Proof. Suppose that (7. 2) holds for $m = 1$. We decompose T^m into the direct product $T^1{}_1 \times T^1{}_2 \times \cdots \times T^1{}_m$ of m circle groups. By our supposition, for each j, there exists a rational element h_j of $\boldsymbol{H}$ such that ih_j generates the circle group $T^1{}_j$. The elements $h_1, h_2, \cdots, h_m$ obviously have all the required properties. Thus to prove (7. 2), it suffices to prove it for $m = 1$.

Let G^* be the linear adjoint group of $\boldsymbol{G}$, and $\pi: G \to G^*$ be the natural homomorphism. The image $\pi(T^1)$ is evidently also a circle group. Let x be an element of $\boldsymbol{G}$ which generates T^1 in G. Then the same element x will generate $\pi(T^1)$ in G^*. Thus $\pi(T^1)$ consists of all the linear transformations $\{\exp t\alpha(x) : t = \text{real numbers}\}$ where α denotes the adjoint representation of $\boldsymbol{G}$. Since $\pi(T^1)$ is a circle group, there exists a real number t_0 different from zero such that $\exp t_0\alpha(x)$ is the identity. It follows that the characteristic root of $t_0\alpha(x)$ must be of the form:

$$2n_1\pi i, 2n_2\pi i, 2n_3\pi i, \cdots,$$

where the n's are integers. Taking $h = t_0x/\pi i$, we see immediately that the characteristic roots of $\alpha(h)$ are all integers. Thus h is a rational element of $\boldsymbol{H}$. Moreover, ih evidently generates the group T^1 in G. (7. 2) is hereby proved.

(7.3) *Let $h_1, h_2, \cdots, h_l$ be a rational base of a Cartan subalgebra $\boldsymbol{H}$ of a complex semi-simple Lie algebra $\boldsymbol{G}$. Let $\boldsymbol{H}^0$ denote the set of all linear combinations with real coefficients of $ih_1, \cdots, ih_l$. Suppose that $\boldsymbol{L}$ is a subalgebra of $\boldsymbol{G}$ whose normaliser $\boldsymbol{M}$ in $\boldsymbol{G}$ contains $\boldsymbol{H}$. Then if $\boldsymbol{H}^0 \cap \boldsymbol{L}$ generates a closed subgroup in a group form G of $\boldsymbol{G}$, so does $\boldsymbol{L}$.*

Proof. Let L, M and T^l be the subgroups generated by $\boldsymbol{L}$, $\boldsymbol{M}$ and $\boldsymbol{H}^0$ in G respectively. Then T^l is a toral group of dimension l, and L is a normal subgroup of M. Since M contains a Cartan subalgebra, M is closed in G. For each element x of $\boldsymbol{G}$, let us denote by $g(x)$ the one-parameter subgroup generated by x in G. On account of a theorem of Malcev [7], to show that L is closed in G, it suffices to show that, for each $x \in \boldsymbol{L}$, the closure $\overline{g(x)}$ of $g(x)$ lies on L. If $\overline{g(x)}$ is not compact, then $\overline{g(x)}$ is isomorphic with the additive group of reals, and hence $\overline{g(x)} = g(x) \subset L$. Now assume $\overline{g(x)}$ to be compact. Since $x \in \boldsymbol{L} \subset \boldsymbol{M}$ and M is closed, $\overline{g(x)} \subset M$. From the fact that $\boldsymbol{H}^0 \subset \boldsymbol{H} \subset \boldsymbol{M}$, we know that T^l is a maximal toral subgroup of M. There exists then an element q of M such that $qg(x)q^{-1} \subset T^l$. But L is a normal subgroup of M and $g(x) \subset L$. Therefore, we have

$$qg(x)q^{-1} \subset T^l \cap L.$$

Let T^* be the identity component of $T^l \cap L$. From the connectedness of $g(x)$, $qg(x)q^{-1} \subset T^*$, and hence $q\overline{g(x)}q^{-1} \subset \overline{T}^*$. The subgroup T^*, being generated by $\boldsymbol{H}^0 \cap \boldsymbol{L}$, is closed by hypothesis. Therefore,

$$q\overline{g(x)}q^{-1} \subset \overline{T}^* = T^* \subset T^l \cap L \subset L$$

whence $\overline{g(x)} \subset q^{-1}Lq = L$. Thus L is closed.

(7.4) *Let $\boldsymbol{H}$ be a Cartan subalgebra of $\boldsymbol{G}$, and $\boldsymbol{H}_1$ a rational subalgebra of $\boldsymbol{H}$. Suppose that $h_1, h_2, \cdots, h_a, h_{a+1}, \cdots, h_l$ form a rational base of $\boldsymbol{H}$ such that $h_1, h_2, \cdots, h_a$ span $\boldsymbol{H}_1$. Then the linear subspace $\boldsymbol{L}$ spanned by*

$$\{\boldsymbol{L} \cap \boldsymbol{H}, e_\beta, e_\rho : \beta \in \Theta, \rho \in \Theta'\}$$

forms a subalgebra of $\boldsymbol{G}$ where Θ denotes the totality of roots β of G such that $[e_\beta, \boldsymbol{H}_1] = 0$, and Θ' the totality of positive roots with respect to the rational base $h_1, h_2, \cdots, h_l$.

Proof. Firstly, we see at once that

$$[\boldsymbol{L} \cap \boldsymbol{H}, \boldsymbol{L}] \subset \boldsymbol{L}, \qquad [e_\beta, e_\gamma] \subset \boldsymbol{L}, \qquad [e_\rho, e_\tau] \in \boldsymbol{L}$$

whenever $\beta, \gamma \in \Theta$ and $\rho, \tau \in \Theta'$. Thus to prove (7.4), it is sufficient to show

that $[e_\beta, e_\rho] \in \boldsymbol{L}$ for $\beta \in \Theta$ and $\rho \in \Theta'$, $\rho \notin \Theta$. For this purpose, let us denote by $\lambda_1, \cdots, \lambda_l$ the coordinates in $\boldsymbol{H}$ with respect to the rational base $h_1, \cdots, h_l$. We write the roots β, ρ as linear combinations of the λ's:

$$\beta = r_1\lambda_1 + \cdots + r_a\lambda_a + r_{a+1}\lambda_{a+1} + \cdots + r_l\lambda_l,$$

$$\rho = s_1\lambda_1 + \cdots + s_a\lambda_a + s_{a+1}\lambda_{a+1} + \cdots + s_l\lambda_l.$$

Since $[e_\beta, \boldsymbol{H}_1] = 0$, $\beta(h') = 0$ for all h' in $\boldsymbol{H}_1$. It follows then $r_1 = r_2 = \cdots = r_a = 0$. Let j be the least integer such that $s_j \neq 0$. The root ρ is positive so that $s_j > 0$. If $j > a$, then $\rho(h_1) = \rho(h_2) = \cdots = \rho(h_a) = 0$, and then $[e_\rho, \boldsymbol{H}_1] = 0$ which contradicts the fact $\rho \notin \Theta$. Therefore, $j \leq a$, and we have

$$\beta + \rho = s_j\lambda_j + \cdots + s_a\lambda_a + (s_{a+1} + r_{a+1})\lambda_{a+1} + \cdots + (s_l + r_l)\lambda_l.$$

Since $s_j > 0$, $\beta + \rho$ is always positive in our sense. The commutator $[e_\beta, e_\rho]$, being either zero or a multiple of the positive root vector $e_{\beta+\rho}$, must belong to $\boldsymbol{L}$. This proves (7.4).

8. Proof of Theorem II. By hypothesis, X is a C-subgroup of K. We consider the subgroups T_0, T_1 and U as given in (7.1). Then

$$\text{rank } (K) = \dim. T_1 + \text{rank } (U), \qquad \text{rank } (X) = \dim. T_0 + \text{rank } (U).$$

Taking account of the known facts that the difference between the dimension and the rank of a compact Lie group is always even, we obtain immediately

$$\dim. T_1 - \dim. T_0 = \text{rank } (K) - \text{rank } (X) \equiv \dim. K - \dim. X \pmod 2.$$

Since $\dim. K - \dim. X = \dim. (K/X)$ is even, it follows that

$$\dim. T_1 - \dim. T_0 \equiv 0 \pmod 2. \tag{8.1}$$

Let $\boldsymbol{K}^0$, $\boldsymbol{X}^0$, $\boldsymbol{T}^0{}_1$, $\boldsymbol{T}^0{}_0$ be the Lie algebras (real Lie algebras) of K, X, T_1, T_0 respectively. We consider the complex form $\boldsymbol{G}$ of $\boldsymbol{K}^0$, and choose a Cartan subalgebra $\boldsymbol{H}$ of $\boldsymbol{G}$. Denoting by e_ρ the normalised root vectors of $\boldsymbol{G}$ [13], we can imbed $\boldsymbol{K}^0$ in $\boldsymbol{G}$ as the totality of linear combinations with real coefficients of the following vectors:

$$\{i(e_\rho + e_{-\rho}), e_\rho - e_{-\rho}, i[e_\rho, e_{-\rho}] : \rho = \text{roots of } \boldsymbol{G}\}.$$

Let G be the simply-connected complex Lie group with $\boldsymbol{G}$ as its Lie algebra. Then K is identified with the subgroup generated by $\boldsymbol{K}^0$ in G. Let $\boldsymbol{H}^0$ be the space spanned, with real coefficients, by $\{i[e_\rho, e_{-\rho}] : \rho = \text{roots of } \boldsymbol{G}\}$. This $\boldsymbol{H}^0$ is a maximal abelian subalgebra of $\boldsymbol{K}^0$. Since maximal abelian subalgebras of $\boldsymbol{K}^0$ are conjugate, we can assume $\boldsymbol{T}^0{}_1 \subset \boldsymbol{H}^0$. Since T_1, T_0 are

toral subgroups and $\boldsymbol{T}^0_0 \subset \boldsymbol{T}^0_1 \subset \boldsymbol{H}^0 \subset \boldsymbol{H}$, we know from (7.2) that there exist linearly independent rational elements $h_1, h_2, \cdots, h_b, h_{b+1}, \cdots, h_a$ such that (i) $ih_1, \cdots, ih_b, ih_{b+1}, \cdots, ih_a$ span $\boldsymbol{T}^0_1$, and (ii) $ih_{b+1}, ih_{b+2}, \cdots, ih_a$ span $\boldsymbol{T}^0_0$ where $a = \dim. T_1$, $a - b = \dim. T_0$. From (8.1), it follows

$$b \equiv 0 \pmod 2. \tag{8.2}$$

From (7.1), U is the semi-simple part of the centraliser of T_1 in K. Thus its Lie algebra $\boldsymbol{U}^0$ is the semi-simple part of the centraliser of $\boldsymbol{T}^0_1$ in $\boldsymbol{K}^0$. A direct calculation tells us that $\boldsymbol{U}^0$ is spanned, with real coefficients, by

$$\{i(e_\beta + e_{-\beta}), e_\beta - e_{-\beta}, i[e_\beta, e_{-\beta}] : \beta \in \Theta\}$$

where Θ denotes the totality of roots β of $\boldsymbol{G}$ such that

$$\beta(h_1) = \beta(h_2) = \cdots = \beta(h_a) = 0.$$

Let $\boldsymbol{H}_1$ be the complex subspace spanned by $h_1, h_2, \cdots, h_a$. This is a rational subalgebra of $\boldsymbol{H}$. We can also regard Θ as the totality of roots β of $\boldsymbol{G}$ such that $[e_\beta, \boldsymbol{H}_1] = 0$. Let l' be the rank of U. Then from (7.1)

$$l' = \text{rank } (K) - \dim. T_1 = l - a. \tag{8.3}$$

Among the vectors $[e_\beta, e_{-\beta}]$, $\beta \in \Theta$, there are exactly l' linearly independent ones because $\{i[e_\beta, e_{-\beta}] : \beta \in \Theta\}$ span a maximal abelian subalgebra of $\boldsymbol{U}^0$. On account of (8.3), we can denote these l' linearly independent elements by $h_{a+1}, h_{a+2}, \cdots, h_l$. Since e_β are normalized root vectors, these h's, being of the form $[e_\beta, e_{-\beta}]$, are rational elements of $\boldsymbol{H}$. Moreover, since $\boldsymbol{T}^0_1 \cap \boldsymbol{U}^0 = 0$, the elements

$$h_1, \cdots, h_b, h_{b+1}, \cdots, h_a, h_{a+1}, \cdots, h_l$$

are linearly independent. Thus they form a rational base of $\boldsymbol{H}$.

We know from (7.1) that X is, locally, the direct product $T_0 \times U$. Thus $\boldsymbol{X}^0 = \boldsymbol{T}^0_0 + \boldsymbol{U}^0$, and thus $\boldsymbol{X}^0$ is spanned, with real coefficients, by

$$\{ih_{b+1}, \cdots, ih_a, ih_{a+1}, \cdots, ih_l, i(e_\beta + e_{-\beta}), e_\beta - e_{-\beta} : \beta \in \Theta\}.$$

Let c be a non-zero real number. On account of (8.2), we can construct $u = b/2$ vectors of the following form:

$$k_1 = ih_1 + ch_2, k_2 = ih_3 + ch_4, \cdots, k_u = ih_{b-1} + ch_b.$$

Now let us consider the complex linear subspace $\boldsymbol{L}_c$ spanned by

$$\{k_1, \cdots, k_u, h_{b+1}, \cdots, h_l, e_\beta, e_\rho : \beta \in \Theta, \rho \in \Theta'\}$$

where Θ' denotes the totality of positive roots of $\boldsymbol{G}$ with respect to the rational

OHIO UNIVERSITY
LIBRARY

base $h_1, h_2, \cdots, h_l$ of $\boldsymbol{H}$. From (7.4), $\boldsymbol{L}_c$ forms a subalgebra of $\boldsymbol{G}$. Moreover, it can be easily seen that

$$\boldsymbol{L}_c \cap \boldsymbol{K}^0 = \boldsymbol{X}^0. \tag{8.4}$$

Firstly, we want to show that $\boldsymbol{L}_c$ generates a closed subgroup in G. For this purpose, let us denote by $\boldsymbol{H}^0$ the totality of linear combinations with real coefficients of $ih_1, ih_2, \cdots, ih_l$. Taking account of the fact $c \neq 0$, we find immediately

$$\boldsymbol{L}_c \cap \boldsymbol{H}^0 = \boldsymbol{H} \cap \boldsymbol{X}^0.$$

Since both $\boldsymbol{H}$ and $\boldsymbol{X}^0$ generate a closed subgroup in G, so does their intersection $\boldsymbol{H} \cap \boldsymbol{X}^0$. Thus $\boldsymbol{L}_c \cap \boldsymbol{H}^0$ generates a closed subgroup in G. The normaliser of $\boldsymbol{L}_c$ in $\boldsymbol{G}$ is evidently $\boldsymbol{H} + \boldsymbol{L}_c$ which contains $\boldsymbol{H}$. (7.3) then tells us that $\boldsymbol{L}_c$ generates a closed subgroup $\boldsymbol{L}_c$ in G.

Let $2t$ be the number of roots of $\boldsymbol{G}$, $2t'$ the number of roots in Θ. Then Θ' contains exactly t roots, and

$$\dim. G = 2\dim. \boldsymbol{G} = 2(2t + l), \dim. K = 2t + l.$$

Moreover, Θ and Θ' have $t - t'$ roots in common. By a counting of linearly independent vectors, we find that

$$\begin{aligned}\dim. L_c &= 2\dim. \boldsymbol{L}_c \\ &= 2(b/2 + l - b + 2t' + t - t') = 2l - b + 2t + 2t' \\ \dim. X &= \dim. X^0 = l - b + 2t'.\end{aligned}$$

It follows then that

$$\dim. G - \dim. L_c = \dim. K - \dim. X. \tag{8.5}$$

Now consider G as a transformation group of G/L in the natural manner. Let O be an orbit of K on G/L_c. Then O is homeomorphic with the coset space $K/(K \cap L_c)$. From (8.4), X is the identity component of $L_c \cap K$, and thus (8.5) tells us that

$$\begin{aligned}\dim. O &= \dim. K - \dim. (L_c \cap K) \\ &= \dim. K - \dim. X = \dim. G - \dim. L_c = \dim. (G/L_c).\end{aligned}$$

The orbit O, being a closed manifold, cannot be imbedded as a proper subset of a manifold of the same dimension. Thus $O = G/L_c$ and thus G/L_c is compact. Since G is simply-connected, and L_c is connected, G/L_c must be simply-connected. From the fact that $K/(K \cap L_c) = O = G/L_c$ is simply-connected, we know that $K \cap L_c$ is connected whence $X = K \cap L_c$. There-

fore, $K/X = O$ is homeomorphic with the complex coset space G/L_c. Theorem II is thus proved.

In the above proof of Theorem II, we have eventually shown that there are infinitely many L_c's serving our purposes. This fact will be used later on. We shall state it precisely as follows:

(8. 6) *Let K be a simply-connected, compact semi-simple Lie group, and X a C-subgroup of K. Suppose that K/X is of even dimension and has vanishing Euler characteristic. Let G be the complex semi-simple Lie group in which K is imbedded as a maximal compact subgroup. Then, to each real number c different from zero, there corresponds a connected complex subgroup L_c of G such that* (i) *G/L_c is homeomorphic with K/X and $X = L_c \cap G$,* (ii) *$L_c \neq L_{c'}$ whenever $c \neq c'$,* (iii) *all the L_c have the same normaliser N whose Lie algebra $\boldsymbol{N}$ contains a Cartan subalgebra of the complex Lie algebra of G.*

Proof. Here, all the notations used in the proof of Theorem II will be adopted. We recall that $\boldsymbol{L}_c$ is spanned by the vectors:

$$\{ih_1 + ch_2, \cdots, ih_{b-1} - ch_b, h_{b+1}, \cdots, h_l, e_\beta, e_\rho : \beta \varepsilon \Theta, \rho \varepsilon \Theta'\}$$

and L_c is the subgroup generated by L_c in G. Connectedness of L_c and property (i) have already been proved in Theorem II. Now by hypothesis, the Euler characteristic of K/X vanishes. Thus rank $(X) <$ rank (K), and then

$$b = \text{dim. } T_1 - \text{dim. } T_0 = \text{rank } (K) - \text{rank } (X) > 0.$$

Hence $\boldsymbol{L}_c \neq \boldsymbol{L}_{c'}$ whenever $c \neq c'$. The groups L_c and $L_{c'}$ they generate must be different. This proves property (ii). As for property (iii), we consider the normaliser of $\boldsymbol{L}_c$ in $\boldsymbol{G}$. It can be easily seen that $\boldsymbol{N}$ is the linear subspaces spanned by

$$\{\boldsymbol{H}, e_\beta, e_\rho : \beta \varepsilon \Theta, \rho \varepsilon \Theta'\}$$

which is independent of the parameter c. Moreover, this $\boldsymbol{N}$ evidently contains the Cartan subalgebra $\boldsymbol{H}$ of $\boldsymbol{G}$. Thus the subgroup N generated by $\boldsymbol{N}$ in G is the common normaliser of all the L_c. This completes the proof.

In what follows, we need that G is effective on G/L_c. Then K is, in general, not simply-connected. Proposition (8. 6) has to be modified in order to suit our purpose.

(8. 7) *Let K be a compact semi-simple Lie group, and X a C-subgroup of K. Suppose that K/X is simply-connected and even-dimensional and has*

vanishing Euler characteristic. Let G be the complex semi-simple Lie group in which K is imbedded as a maximal compact subgroup. If K acts effectively on K/X, then, in G, there is a collection $\{L_c\colon c = \text{real number}, c \neq 0\}$ of complex subgroups having the properties (i), (ii), (iii) *in* (8.6) *as well as the property* (iv) *G acts effectively on G/L_c.*

Proof. Let $\tilde{K}$ be the universal covering group of K, and $\tilde{X}$ the subgroup of $\tilde{K}$ which covers X. Since $\tilde{K}/\tilde{X} = K/X$ is simply-connected, $\tilde{X}$ is connected. We can easily see that $\tilde{X}$ is a C-subgroup of $\tilde{K}$. Let $\tilde{G}$ be the complex semi-simple Lie group in which $\tilde{K}$ is imbedded as a maximal compact subgroup. By (8.6), there is a collection $\{\tilde{L}_c\}$ of complex subgroups of $\tilde{G}$ having the properties (i), (ii), (iii) with respect to $\tilde{K}$, $\tilde{X}$, G. Let $\pi\colon \tilde{G} \to G$ be the projection. Up to an automorphism of G, we can assume $\pi(\tilde{K}) = K$, $\pi(\tilde{X}) = X$. The subgroups $L_c = \pi(\tilde{L}_c)$ evidently have the properties (i), (ii), (iii) with respect to K, X, G. It remains only to show that G acts effectively on G/L_c, or what is the same, that L_c contains no normal subgroup of G except the identity. Let Q be the maximal normal subgroup of G contained in L_c. Then $K \cap Q \subset L_c \cap K = X$. From the effectiveness of K on K/X, we know that $K \cap Q$ contains only the identity. A simple argument on Lie algebras tells us that dim. $Q =$?dim. $(K \cap Q)$, and hence Q must be discrete. Thus Q belongs to the center of G which belongs to K. It follows that Q contains only the identity. The effectiveness of G on G/L_c is hereby established.

9. Two propositions. In order to study the group of all analytic homeomorphisms of a C-space, we shall first prove two rather simple propositions:

(9.1) *Let $W = B/L_1$ be a left coset space of a Lie group B. Suppose G to be a closed subgroup of B transitive over W. If B is effective on W, then the Lie algebra $\boldsymbol{L}_1$ of L_1 cannot contain any non-zero subspace $\boldsymbol{Q}$ with $[\boldsymbol{G}, \boldsymbol{Q}] \subset \boldsymbol{Q}$, where $\boldsymbol{G}$ denotes the Lie algebra of G.*

Proof. Let $\boldsymbol{V}$ be the least subalgebra which contains $\boldsymbol{Q}$. Evidently, $\boldsymbol{V} \subset \boldsymbol{L}_1$, $[\boldsymbol{G}, \boldsymbol{V}] \subset \boldsymbol{V}$. Let V be the subgroup generated by $\boldsymbol{V}$ in B. Then $V \subset L_1$, and G is contained in the normaliser of V. Since G is transitive on W, each element of W takes the form gL_1 where $g \in G$. Let $v \in V$. We have $vg = g(g^{-1}vg) \in gV \subset gL_1$, and thus $vgL_1 = gL_1$ for each g of G. This tells us that v induces the identity transformations on W. By hypothesis, B is effective so that V contains only the identity. Therefore, the subalgebra $\boldsymbol{V}$ as well as the linear subspace $\boldsymbol{Q}$ must be zero. (9.1) is proved.

(9.2) *Let ϕ be a non-trivial irreducible representation of a complex semi-simple Lie algebra $\boldsymbol{G}$ over a vector space V, and V^* a subspace of V invariant under $\phi(e_\rho)$ for all the positive root vectors e_ρ of $\boldsymbol{G}$. If, moreover, V^* contains all the weight vectors with non-vanishing weight, then $V = V^*$.*

Proof. We note that we have a fixed Cartan subalgebra $\boldsymbol{H}$ based on which the weight vectors, weight spaces and weights are defined. Moreover, we have a fixed rational base of $\boldsymbol{H}$ by which the ordering of weights and the ordering of roots of $\boldsymbol{G}$ are determined. Since ϕ is non-trivial, the minimum weight Λ of ϕ must be negative. Take a vector v with weight Λ. Then each vector v' of the form:

$$v' = E_{\alpha_1}E_{\alpha_2}\cdots E_{\alpha_s}(v)$$

is a weight vector with weight equal to $\Lambda + \alpha_1 + \alpha_2 + \cdots + \alpha_s$, where $E_\alpha = \phi(e_\alpha)$, e_α denoting the root vector corresponding to the root α. Since ϕ is irreducible, all the vectors v' of the above form span the entire space V. Thus to prove our proposition, it suffices to show that each v' belongs to V^*.

If $\Lambda + \alpha_1 + \alpha_2 + \cdots + \alpha_s \neq 0$, then $v' \varepsilon V^*$ by hypothesis. Now assume $\Lambda + \alpha_1 + \cdots + \alpha_s = 0$. We shall call s the length of the vector v', and use induction on s. If the length s is zero, then $v' = v \varepsilon V^*$. Suppose $v' \varepsilon V^*$ whenever the length is less than s. We consider a v' of length s. Since $\Lambda + \alpha_1 + \cdots + \alpha_s = 0$ and Λ is negative, one of the α's must be positive. Let j be the least integer such that α_j is positive. We can easily verify that

$$v' = E_{\alpha_1}E_{\alpha_2}\cdots E_{\alpha_s}(v) = E_{\alpha_j}E_{\alpha_1}\cdots E_{\alpha_{j-1}}E_{\alpha_{j+1}}\cdots E_{\alpha_s}(v) + \sum_{i=1}^{j-1} F_i(v)$$

where

$$F_i = E_{\alpha_1}E_{\alpha_2}\cdots E_{\alpha_{i-1}}[E_{\alpha_i}E_{\alpha_j}]E_{\alpha_{i+1}}\cdots E_{\alpha_{j-1}}E_{\alpha_{j+1}}\cdots E_{\alpha_s}.$$

Since $\Lambda + \alpha_1 + \cdots + \alpha_{j-1} + \alpha_{j+1} + \cdots + \alpha_s = -\alpha_j \neq 0$,

$$v'' = E_{\alpha_1}E_{\alpha_2}\cdots E_{\alpha_{j-1}}E_{\alpha_{j+1}}\cdots E_{\alpha_s}(v) \varepsilon V^*.$$

From the invariance of V^* under $E_{\alpha_j} = \phi(e_{\alpha_j})$, we know that $E_{\alpha_j}(v'') \varepsilon V^*$. As for F_i, there are three possibilities. If $\alpha_i + \alpha_j$ is not a root, then $F_i = 0$; if $\alpha_i + \alpha_j = 0$, then $F_i(v)$ is a vector of length $s-2$; and if $\alpha_i + \alpha_j$ is a root of $\boldsymbol{G}$, then $F_i(v)$ is a multiple of a vector of length $s-1$. Thus, by induction hypothesis, $F_i(v)$ always belongs to V^*. It follows then that $v' = E_{\alpha_j}(v'') + \sum_{i=1}^{j-1} F_i(v)$ belongs to V^*. The induction is completed and (9.2) is proved.

10. Group of all analytic homomorphisms. Let B be the group of all analytic homeomorphisms of a C-space. From a theorem of Bochner-Montgomery, it follows that B is a complex Lie group. In the present section, we shall show that the radical of B coincides with the center of B.

THEOREM III. *Let B be the identity component of the group of all analytic homeomorphisms of a C-space W. Then B is, locally, the direct product of a complex semi-simple Lie group and a complex vector group.*

Proof. From § 3, we know that B is a complex Lie group and that the isotropic subgroup L_1 of B is a complex subgroup of B. Since $W = B/L_1$ is simply-connected, L_1 must be connected. Let $B = GR$ be a Levi-decomposition of B where G denotes a maximal semi-simple subgroup of B and R the radical. Choose a maximal connected solvable subgroup M of the intersection $L = L_1 \cap G$. This M is contained in a certain maximal connected solvable subgroup, say M_1, of L_1 Evidently, M is the identity component of $M_1 \cap G$. We shall first show

$$\text{dim. } M_1 - \text{dim. } M = \text{dim. } R. \tag{10.1}$$

For this purpose, we observe that B/M_1 is a fibre bundle over B/L_1 with L_1/M_1 as the fibre [11], p. 30. The base space B/L_1, being homeomorphic with W, is compact and simply-connected. By using first (3. 5) and then (3. 1), we see that L_1/M_1 is compact. Thus the bundle space B/M_1 is compact. Moreover, (3. 5) tells us that $\pi_1(L_1/M_1)$ is finite. From this fact and the simply-connectedness of B/L_1, it follows that the fundamental group $\pi_1(B/M_1)$ is finite. Proposition (2. 3) then tells us that the maximal semi-simple subgroup G is transitive over B/M_1, and hence $B/M_1 = G/(G \cap M_1)$. We have, therefore,

$$\begin{aligned}\text{dim. } G + \text{dim. } R - \text{dim. } M_1 &= \text{dim. } B - \text{dim. } M_1 \\ &= \text{dim. } G - \text{dim. } (M_1 \cap G) = \text{dim. } G - \text{dim. } M\end{aligned}$$

which gives (10. 1).

Let $\boldsymbol{B}, \boldsymbol{G}, \boldsymbol{R}, \boldsymbol{L}_1, \boldsymbol{L}, \boldsymbol{M}_1, \boldsymbol{M}$ be the Lie algebras of B, G, R, L_1, L, M_1, M respectively. All of them are complex Lie algebras. Passing the relations among these groups to their respective algebras, we obtain (a) $\boldsymbol{B} = \boldsymbol{G} + \boldsymbol{R}$ is a Levi-decomposition of the complex Lie algebra $\boldsymbol{B}$, (b) $\boldsymbol{M}$ is a maximal solvable subalgebra of $\boldsymbol{L}$, (c) $\boldsymbol{M}_1$ is a maximal solvable subalgebra of $\boldsymbol{L}_1$, (d) $\boldsymbol{L} = \boldsymbol{L}_1 \cap \boldsymbol{G}$, $\boldsymbol{M} = \boldsymbol{M}_1 \cap \boldsymbol{G}$. Now choose a maximal solvable subalgebra $\boldsymbol{S}_1$ of $\boldsymbol{B}$ such that $\boldsymbol{M}_1 \subset \boldsymbol{S}_1$. Obviously, $\boldsymbol{S}_1$ can be written in the form: $\boldsymbol{S}_1 = \boldsymbol{S} + \boldsymbol{R}$ where $\boldsymbol{S}$ denotes a maximal solvable subalgebra of $\boldsymbol{G}$. Thus

$$\boldsymbol{M} = \boldsymbol{M}_1 \cap \boldsymbol{G} = (\boldsymbol{M}_1 \cap \boldsymbol{S}_1) \cap \boldsymbol{G} = \boldsymbol{M}_1 \cap (\boldsymbol{S}_1 \cap \boldsymbol{G}) = \boldsymbol{M}_1 \cap \boldsymbol{S}.$$

From the theory of vector spaces,

$$\dim.\ (\boldsymbol{S}+\boldsymbol{M}_1)+\dim.\ (\boldsymbol{M}_1\cap\boldsymbol{S})=\dim.\ \boldsymbol{S}+\dim.\ \boldsymbol{M}_1,$$

or what is the same,

$$\dim.\ (\boldsymbol{S}+\boldsymbol{M}_1)=\dim.\ \boldsymbol{S}+\dim.\ \boldsymbol{M}_1-\dim.\ \boldsymbol{M}.$$

On the other hand, we have from (10. 1) that $\dim.\ \boldsymbol{M}_1-\dim.\ \boldsymbol{M}=\dim.\ \boldsymbol{R}$. Therefore, $\dim.(\boldsymbol{S}+\boldsymbol{M}_1)=\dim.\ \boldsymbol{S}+\dim.\ \boldsymbol{R}$. But $\dim.\ \boldsymbol{S}_1=\dim.\ \boldsymbol{S}+\dim.\ \boldsymbol{R}$ and $\boldsymbol{S}+\boldsymbol{M}_1\subset\boldsymbol{S}_1$. It follows then

$$\boldsymbol{R}\subset\boldsymbol{S}+\boldsymbol{R}=\boldsymbol{S}_1=\boldsymbol{S}+\boldsymbol{M}_1. \tag{10. 2}$$

Now let us consider $\boldsymbol{S}$. Since it is a maximal solvable subalgebra of $\boldsymbol{G}$, there exists a Cartan subalgebra $\boldsymbol{H}$ of $\boldsymbol{G}$ and a rational base $h_1, h_2, \cdots, h_l$ of $\boldsymbol{H}$ such that $\boldsymbol{S}$ is spanned by $\boldsymbol{H}$ together with all the positive root vectors of $\boldsymbol{G}$. Since $W=B/L_1$ is a C-space, (2. 3) tells us that G is transitive over W, and $W=G/(L_1\cap G)=G/L$. Taking account of the fact $\boldsymbol{M}\subset\boldsymbol{S}$, we know from the first two paragraphs of § 5 that (i) $\boldsymbol{L}\cap\boldsymbol{H}$ is a J-subalgebra of $\boldsymbol{H}$, (ii) $\boldsymbol{L}$ is spanned by $\boldsymbol{L}\cap\boldsymbol{H}$ and some root vectors, and (iii) all the positive root vectors belong to $\boldsymbol{L}$. From (iii), it follows immediately

$$[\boldsymbol{S},\boldsymbol{S}]\subset\boldsymbol{L}\subset\boldsymbol{L}_1. \tag{10. 3}$$

Let ϕ be the representation of $\boldsymbol{G}$ over the radical $\boldsymbol{R}$ defined by

$$\phi(g)(w)=[g,w],\ g\in\boldsymbol{G},\ w\in\boldsymbol{R}.$$

This is a well-defined representation as $\boldsymbol{R}$ is the radical. We consider a non-trivial irreducible component ϕ_1 of ϕ, and denote by R' the subspace of $\boldsymbol{R}$ corresponding to ϕ_1. Then R' is decomposed into weight spaces:

$$R'=R_{\Lambda_1}+R_{\Lambda_2}+\cdots$$

where Λ_j denote the weights. Let w be a vector of R_{Λ_j} with $\Lambda_j\neq 0$. From (10. 2), we know that w can be written in the form: $w=x+y$, $x\in\boldsymbol{S}$, $y\in\boldsymbol{M}_1$. Since $\{h:\ h\in\boldsymbol{H},\ \Lambda_j(h)=0\}$ is a rational hyperplane and $\boldsymbol{L}\cap\boldsymbol{H}$ is a J-subalgebra, there exists $h_0\in\boldsymbol{H}\cap\boldsymbol{L}$ such that $\Lambda_j(h_0)\neq 0$. Thus

$$\Lambda_j(h_0)w=\phi(h_0)(w)=[h_0,x]+[h_0,y],$$

and thus

$$w=([h_0,x]+[h_0,y])/\Lambda_j(h_0).$$

Both h_0 and y belong to $\boldsymbol{L}_1$ so that $[h_0,y]\in\boldsymbol{L}_1$. As for $[h_0,x]$, we have from (10. 3)

$$[h_0,x]\in[\boldsymbol{H},\boldsymbol{S}]\subset[\boldsymbol{S},\boldsymbol{S}]\subset\boldsymbol{L}_1.$$

Therefore, $w \in \boldsymbol{L}_1$, and then $w \in \boldsymbol{L}_1 \cap \boldsymbol{R}'$. Since this holds for all weight vectors with non-vanishing weight, we know that the subspace $\boldsymbol{L} \cap \boldsymbol{R}'$ contains all the weight vectors with non-vanishing weight. Moreover, for each positive root vector e_ρ of G, $e_\rho \in \boldsymbol{L}_1$ so that

$$\phi(e_\rho)(\boldsymbol{L}_1 \cap \boldsymbol{R}') = [e_\rho, \boldsymbol{L}_1 \cap \boldsymbol{R}'] \subset \boldsymbol{L}_1 \cap \boldsymbol{R}'.$$

In other words, the subspace $\boldsymbol{L}_1 \cap \boldsymbol{R}'$ of $\boldsymbol{R}'$ is invariant under $\phi(e_\rho)$. Proposition (9. 2) then tells us that $\boldsymbol{R}' = \boldsymbol{L}_1 \cap \boldsymbol{R}'$, or what is the same, $\boldsymbol{R}' \subset \boldsymbol{L}_1$. Taking account of the fact $[\boldsymbol{G}, \boldsymbol{R}'] \subset \boldsymbol{R}'$, we see from (9. 1) that $R' = 0$ because of the effectiveness of B on W. Therefore, each irreducible component of ϕ must be trivial. In other words, ϕ is trivial. We have then $[\boldsymbol{G}, \boldsymbol{R}] = 0$. This shows that the Levi-decomposition of $\boldsymbol{B}$ is a *direct* sum.

It remains for us only to show that the radical $\boldsymbol{R}$ is abelian. Let w, w' be two vectors in $\boldsymbol{R}$. On account of (10. 2), we can write them in the form:

$$w = x + y, \qquad w' = x' + y', \qquad x, x' \in \boldsymbol{S} \subset \boldsymbol{G}, \qquad y, y' \in \boldsymbol{M}_1 \subset \boldsymbol{L}_1.$$

Thus

$$[y, y'] = [w - x, w' - x'] = [w, w'] + [x, x'],$$

and thus

$$[w, w'] = [y, y'] - [x, x'].$$

The right hand side belongs to $\boldsymbol{L}_1$ on account of (10. 3) while the left hand side belongs to $\boldsymbol{R}$. Hence $[w, w'] \in \boldsymbol{L}_1 \cap \boldsymbol{R}$. From (9. 1) and the fact $[\boldsymbol{G}, \boldsymbol{R}] = 0$, we know that $\boldsymbol{L}_1 \cap \boldsymbol{R} = 0$, whence $[w, w'] = 0$. $\boldsymbol{R}$ is therefore abelian and Theorem III is proved.

11. Homogeneous complex structures over a given manifold. Let M be a manifold over which two complex structures $\mathfrak{A}$ and $\mathfrak{A}'$ are defined. These two structures are called *equivalent* if the two complex manifolds $(M, \mathfrak{A})$ and $(M, \mathfrak{A}')$ are analytically homeomorphic. It is the aim of the present section to study the number of inequivalent homogeneous complex structures over a given manifold.

THEOREM IV. *Let M be a simply-connected closed manifold with non-vanishing Euler characteristic. Then M has at most a finite number of inequivalent homogeneous complex structures.*

Proof. On account of (2. 3), it is sufficient to show that there are only at most a finite number of analytically non-homeomorphic coset spaces G/L whose underlying spaces are homeomorphic with M where G denotes a complex

semi-simple Lie group. Let dim. $M = n$. Since the maximal compact subgroup $K(G)$ of G is transitive over M, we have dim. $K(G) \leq n(n+1)/2$. Thus dim. $G \leq n(n+1)$ and thus up to isomorphisms, there are only a finite number of possible G's. Now fix the complex semi-simple Lie group G. Since the Euler characteristic of M is different from zero, we can see without difficulty that the Lie algebra $\boldsymbol{L}$ of L contains a Cartan subalgebra $\boldsymbol{H}$ of the complex Lie algebra $\boldsymbol{G}$ of G. Thus $\boldsymbol{L}$ is spanned by $\boldsymbol{H}$ and some root vectors of $\boldsymbol{G}$. It follows then that, up to inner automorphisms of $\boldsymbol{G}$, such subalgebras $\boldsymbol{L}$ are finite in number. Thus, up to automorphisms of G, we have only a finite number of such subgroups L. Hence, the number of different coset spaces G/L homeomorphic with M is finite. This proves our theorem.

When the Euler characteristic of M vanishes, the situation is entirely different. We shall see, in this case, that M has either no homogeneous complex structure or non-countably many inequivalent ones.

(11. 1) *Let $W = G/L$, $W' = G'/L'$ be two C-spaces where G, G' denote respectively the maximal connected semi-simple subgroups of the groups B, B' of all analytic homeomorphisms of W, W'. If W and W' are analytic homeomorphic, then there is an isomorphism θ of G onto G' such that $\theta(L) = L'$.*

Proof. From Theorem III, we know that B as well as B' has only one maximal connected semi-simple subgroups. From this fact, Proposition (11. 1) follows immediately.

(11. 2) *Let N be a closed subgroup of a complex semi-simple Lie group G such that N contains a maximal toral subgroup T of G. Let A be the group of all automorphisms of G which leave N invariant, and A_0 the group of all those inner automorphisms induced by elements in N. Then A/A_0 is finite.*

Proof. We first observe that if an automorphism $\beta \in A$ leaves invariant the subgroup T, then its differential $d\beta$ leaves invariant a Cartan subalgebra $\boldsymbol{H}$ of the complex Lie algebra $\boldsymbol{G}$ of G. Therefore, such automorphisms are finite in number. Let us denote them by $\beta_1, \beta_2, \cdots, \beta_s$. We shall show that $A = A_0\beta_1 \cup A_0\beta_2 \cup \cdots \cup A_0\beta_s$. Suppose $\alpha \in A$. Then $\alpha(T) \subset N$. Since T and $\alpha(T)$ are two maximal toral subgroups in N, there exists an inner automorphism γ of N carrying T onto $\alpha(T)$. Since $N \subset G$, we can regard this γ as an inner automorphism of G, and thus $\gamma \in A_0$. The automorphism $\gamma^{-1}\alpha$ leaves invariant T. It must be one of the β's, say β_j. We

have then $\alpha = \gamma\beta_j \in A_0\beta_j$. This tells us that A/A_0 cannot have more than s elements. (11.2) is proved.

(11.3) *Let G be a complex semi-simple Lie group, and $\{L_c\}$ a collection of closed subgroups depending on a non-zero real parameter c. Suppose that* (i) *$L_c \neq L_{c'}$ whenever $c \neq c'$,* (ii) *all the L_c have the same normaliser N in G,* (iii) *the Lie algebra $\mathfrak{N}$ of N contains a Cartan subalgebra $\mathfrak{H}$ of the complex Lie algebra $\mathfrak{G}$ of G. Then, in the collection $\{L_c\}$ of subgroups, there are non-countably many inequivalent ones under automorphisms of G.*

Proof. We divide the collection $\{L_c\}$ of subgroups into equivalent classes by automorphisms of G. To prove (11.3), it is sufficient to show that each equivalent class contains only a finite number of L_c's. Suppose that L_c and $L_{c'}$ are equivalent. Then there exists an automorphism α of G carrying L_c onto $L_{c'}$. Evidently, α will carry the normaliser of L_c to that of $L_{c'}$ or what is the same, $\alpha(N) = N$. On account of property (iii), N contains a maximal toral subgroup of G so that (11.2) can be applied. Using the same notations A, A_0 as in (11.2), we know that there exist elements $\sigma_1, \sigma_2, \cdots, \sigma_m$ of A with the property

$$A = \sigma_1 A_0 \cup \sigma_2 A_0 \cup \cdots \cup \sigma_m A_0.$$

Since $\alpha \in A$, $\alpha = \sigma_j\gamma$ for a certain j and a certain γ of A_0. The group N is the normaliser of L_c, and γ is an inner automorphism of N. Thus $L_{c'} = \alpha(L_c) = \sigma_j\gamma(L_c) = \sigma_j(L_c)$. This tells us that each equivalent class contains at most m different L_c's. Our Proposition (11.3) is hereby proved.

THEOREM V. *Let $\mathfrak{M}$ be a simply-connected closed manifold. If its Euler characteristic vanishes, then it has either no homogeneous complex structure or non-countably many inequivalent ones.*

Proof. Suppose that M has a homogeneous complex structure. Then M is homeomorphic with a coset space K/X where K is a compact semi-simple Lie group and X a C-subgroup of K. As can be easily justified, we can assume that K acts *effectively* on K/X. Of course, there may be various pairs (K, X) with these properties. Now we choose (K, X) so that K is of highest possible dimension. Let G be the complex semi-simple Lie group in which K is imbedded as a maximal compact subgroup. From (8.6), there exist a collection $\{L_c\}$ of closed complex subgroups of G depending on a non-zero real parameter c with the following properties: (i) G/L_c is homeomorphic with K/X for each c, (ii) $L_c \neq L_{c'}$ whenever $c \neq c'$, (iii) all the L_c have the same normaliser N whose Lie algebra contains a Cartan subalgebra of the

Lie algebra of G, (iv) G acts effectively on G/L_c. Evidently, each G/L_c is a C-space homeomorphic with M. Thus, to prove our theorem, it suffices to prove that, in the collection $\{L_c\colon c \neq 0,\ c = \text{real number}\}$ of complex coset spaces, there are non-countably many spaces which are analytically non-homeomorphic with one another.

Let $W_c = G/L_c$, and B be the group of all analytic homeomorphisms of W_c. In the natural manner, each element of G can be regarded as an analytic homeomorphism of W_c. On account of the effectiveness of G on W_c, we can regard G as a subgroup of B. From Theorem III, B has a unique maximal connected semi-simple subgroup G^*. Thus $G \subset G^*$. We know from the proof of Theorem I that W_c is homeomorphic with $K(G^*)/X^*$ where $K(G^*)$ denotes a maximal compact subgroup of G^*, and X^* a C-subgroup of $K(G^*)$. From our choice of our K and X, it follows that $\dim. K(G^*) \leq \dim. K$. Hence $\dim. G^* = 2\dim. K(G^*) \leq 2\dim. K = \dim. G$, and hence $G = G^*$. In other words, G is the maximal connected semi-simple subgroup of the group of all analytic homeomorphisms of W_c. Taking account of this fact, we know from (11. 1) that W_c and $W_{c'}$ are analytically homeomorphic if and only if L_c and $L_{c'}$ are equivalent under automorphisms of G. (11. 3) tells us that there are non-countably many inequivalent L_c's. Therefore, there are non-countably many analytically non-homeomorphic W_c's. The theorem is hereby proved.

12. Determination of C-spaces. To determine all the C-spaces, we shall introduce a class of closed manifolds which are closely related with C-spaces. Let K be a simply-connected compact simple Lie group, and U a semi-simple C-subgroup of K. The coset space K/U is called an *M-manifold*. From the homotopy sequence, it follows immediately that an M-manifold is simply-connected and has a finite second homotopy group. Thus its second Betti number vanishes.

THEOREM VI. *Each C-space W is a fibre decomposition space of a product M of M-manifolds with torus as the fibre* (i. e., *the product M is a torus bundle over W*). *Conversely, suppose $M = M_1 \times M_2 \times \cdots \times M_s$ to be a product of M-manifolds K_j. If M is even-dimensioned, then M has infinitely many inequivalent homogeneous complex structures. Moreover, the resulting complex manifolds are non-kählerian.*

Proof. Let W be a C-space. By Theorem I, it is homeomorphic with a coset space K/X where K is a compact semi-simple Lie group and X a C-

subgroup of K. Let U be the maximal connected semi-simple subgroup of X. Then U is a normal subgroup of X, and X/U is a toral group. From the theory of fibre bundles [11], we know that K/U is a fibre bundle over K/X with X/U as the fibre. Thus W is a fibre decomposition space of K/U with the torus X/U as the fibre. Hence, to prove the first part of Theorem VI, it suffices to prove that K/U is a product of M-manifolds. Firstly, we observe that K can be assumed simply connected, for otherwise, we can take the universal covering group of K in place of K. Thus K is decomposed into a direct product of simply-connected, compact simple Lie groups:

$$K = K_1 \otimes K_2 \otimes \cdots \otimes K_s.$$

By definition of C-subgroups, U is the semi-simple part of the centraliser Q of a toral subgroup of K. Evidently, Q has the same rank as K. It follows then [5]

$$Q = (K_1 \cap Q) \otimes (K_2 \cap Q) \otimes \cdots \otimes (K_s \cap Q).$$

Thus $U = U_1 \otimes U_2 \otimes \cdots \otimes U_s$ where U_i denotes the semi-simple part of $K_i \cap Q$, and thus

$$K/U = K_1/U_1 \times K_2/U_2 \times \cdots \times K_s/U_s. \tag{12.1}$$

We can verify without difficulty that U_i is a C-subgroup of K_i. Moreover, each K_i is a simply-connected, compact simple Lie group. Therefore, K_i/U_i are M-manifolds. (12.1) then proves the first part of our theorem.

As for the second part, we write $M_i = K_i/U_i$ where K_i denotes a simply-connected, compact simple Lie group and U_i a semi-simple C-subgroup of K_i. Then the product space $M = M_1 \times M_2 \times \cdots \times M_s$ is homeomorphic with the coset space

$$(K_1 \otimes K_2 \otimes \cdots \otimes K_s)/(U_1 \otimes U_2 \otimes \cdots \otimes U_s).$$

The product $K_1 \otimes K_2 \otimes \cdots \otimes K_s$ is obviously a simply-connected and compact semi-simple Lie group. Moreover, it is easy to see that $U_1 \otimes U_2 \otimes \cdots \otimes U_s$ is a C-subgroup of $K_1 \otimes K_2 \otimes \cdots \otimes K_s$. The existence of homogeneous complex structure over M then follows from Theorem II. By hypothesis, each M_i is an M-manifold so it has vanishing second Betti number. Thus M also has vanishing second Betti number. It follows that M can never be rendered a kählerian manifold. Besides, we know from the Corollary in § 6 that the Euler characteristic of M must vanish. Theorem V then tells us

that M has infinitely many inequivalent homogeneous complex structures. Theorem VI is hereby proved.

From the above theorem, the study of the topological properties of C-spaces is reduced to that of M-manifolds. By definition, an M-manifold is a coset space K/U where K is a compact simple Lie group and U a semi-simple C-subgroup of K. Thus U is the centraliser Q of a toral subgroup of K. Taking account of the fact that Q has the same rank as K, we can determine all the U's by using known results [5], [12]. Here we shall simply state the results without elaborating.

THEOREM VII. *The M-manifolds $M = K/U$ corresponding to a classical group K are given as follows*:

$$A_l/(A_{i_1} \times A_{i_2} \times \cdots \times A_{i_a}), \qquad 0 \leq i_1 + i_2 + \cdots + i_a + a \leq l-1,$$

$$B_l/(A_{i_1} \times A_{i_2} \times \cdots \times A_{i_a} \times B_b), \quad 0 \leq i_1 + i_2 + \cdots + i_a + a + b \leq l-1$$

$$C_l/(A_{i_1} \times A_{i_2} \times \cdots \times A_{i_a} \times C_c), \quad 0 \leq i_1 + i_2 + \cdots + i_a + a + c \leq l-1$$

$$D_l/(A_{i_1} \times A_{i_2} \times \cdots \times A_{i_a} \times D_d), \quad 0 \leq i_1 + i_2 + \cdots + i_a + a + d \leq l-1$$

where A_j, B_j, C_j, D_j denote, respectively, the compact simple Lie groups of Cartan classes A, B, C, D with rank j.

As for the exceptional groups, we will not enumerate all their semi-simple C-subgroups. Only the maximal ones will be given. In fact, this is sufficient as we can easily prove the following:

Let U, U^* be two semi-simple C-subgroups of K. If $U \subset U^*$ then U is a C-subgroup of U^*.

THEOREM VIII. *All the semi-simple maximal C-subgroups of the exceptional groups are given by the following table*:

Group	maximal semi-simple C-subgroups
G_2	A_1
F_4	$B_3, C_3, A_1 \times A_2$
E_6	$D_5, A_5, A_1 \times A_4, A_1 \times A_2 \times A_2$
E_7	$E_6, A_6, A_1 \times A_5, A_2 \times A_4, A_1 \times A_2 \times A_3$
E_8	$E_7, D_7, A_7, A_1 \times E_6, A_1 \times A_6, A_2 \times D_5, A_3 \times A_4, A_1 \times A_2 \times A_4$

Added in Proof. In connection with the results in this paper, H. Samelson has recently proved that the underlying space of an even-dimensional compact Lie group has always a homogeneous complex structure (to appear in Portugaliae Mathematica).

ALABAMA POLYTECHNIC INSTITUTE,
AUBURN, ALABAMA.

REFERENCES.

[1] S. Bochner and W. T. Martin, *Several complex variables*, Princeton, 1948.

[2] ——— and D. Montgomery, "Groups of differential and real or complex analytic transformations," *Annals of Mathematics*, vol. 46 (1945), pp. 685-694.

[3] ——— and D. Montgomery, "Groups of analytic manifolds," *ibid.*, vol. 48 (1947), pp. 659-669.

[4] A. Borel, "Sur la cohomologie des espaces fibres principaux et des espaces homogènes de groupes de Lie compacts," *ibid.*, vol. 57 (1953), pp. 115-205.

[5] ——— and J. De Siebenthal, "Les sous-groupes fermes de rang maximum des groupes de Lie close," *Comment. Math. Helv.*, vol. 23 (1949), pp. 200-221.

[6] E. Calabi and B. Eckmann, "A class of simply-connected compact complex manifolds which are not algebraic," *Bulletin of the American Mathematical Society*, vol. 57 (1951), p. 533.

[7] A. Malcev, "Subgroups of Lie groups in the large," *Comptes Rendus (Doklady) Acad. Sci. URSS*, Ser., vol. 36 (1942), pp. 5-7.

[8] ———, "On semi-simple subgroups of Lie groups," *Bulletin Acad. Sci. URSS*, Ser. Math. (Izvestiya Akad. Nauk SSSR), vol. 8 (1944), pp. 143-174.

[9] D. Montgomery, "Simply-connected homogeneous spaces," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 467-469.

[10] V. V. Morozoff, "On a nilpotent element in a semi-simple Lie algebra," *Comptes Rendus (Doklady) Akad. Sci. URSS*, vol. 36 (1942), pp. 83-86.

[11] N. E. Steenrod, *The topology of fibre bundles*, Princeton, 1951.

[12] H. C. Wang, "Homogeneous spaces with non-vanishing Euler characteristic," *Annals of Mathematics*, vol. 50 (1949), pp. 925-953.

[13] H. Weyl, "Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare Transformationen, I," *Mathematische Zeitschrift*, vol. 23 (1924), II, III, vol. 24 (1925), pp. 328-395.

INVARIANT AFFINE CONNECTIONS ON HOMOGENEOUS SPACES.*

By KATSUMI NOMIZU.

Introduction. The purpose of the present paper is to clarify and generalize the work of E. Cartan on the relationship between differential geometry and the spaces of Lie groups and the homogeneous spaces. We study the existence and properties of invariant affine connections on homogeneous spaces of certain types, in particular, on symmetric homogeneous spaces. By a symmetric homogeneous space we mean a homogeneous space G/H where G is a Lie group with an involutive automorphism σ and H is the subgroup of fixed point of σ.

Our problem may be explained, roughly, as follows. Let G/H be a homogeneous space of a connected Lie group. It is well known that if the subgroup H is compact, one can introduce a positive definite riemannian metric on G/H which is invariant by G acting on G/H in the natural fashion. The affine connection which is canonically associated to this invariant metric is then an invariant affine connection on G/H. The question arises, therefore, whether there exist invariant affine connections on G/H when H is not necessarily compact. Even when it is compact, there may exist invariant affine connections which do not arise from any metrics. In the case of a symmetric homogeneous space G/H with compact H, the invariant affine connection arising from any invariant metric has certain remarkable properties, first noted by E. Cartan, and serves as an important tool for the investigation of the space. If one attempts to generalize his theory to any symmetric homogeneous space, it is of primary importance to establish the existence of an invariant affine connection on it, having properties similar to those of the riemannian conncetion in Cartan's study.

We shall investigate these questions for a fairly general class of homogeneous spaces, which we call reductive and which includes the symmetric homogeneous spaces. In Chapter I, we shall give a general definition of affine connection, which is due to Koszul, and we shall formulate certain basic concepts in the theory of affine connections in a form convenient for our

---

* This paper was written while the author was a Eugene Higgins Fellow at Columbia University.

purpose. In Chapter II, we shall study the existence and properties of invariant affine connections on reductive homogeneous spaces. We shall see that the problems which arise can be reduced to algebraic questions concerning the Lie algebra of G and that of H. Specifically, we show that there exists a one-to-one correspondence between the set of invariant affine connections on G/H and the set of bilinear functions, called connection functions, which are themselves of simple nature (Theorem 8. 1). In particular, there exist always two important affine connections, called the canonical affine connections of the first and second kind (Theorem 10. 1 and Theorem 10. 2). The well known theorem concerning the holonomy group of a riemannian symmetric homogeneous space is obtained for the canonical affine connection of the second kind on any reductive homogeneous space (Theorem 12. 1).

Chapter III is devoted to a treatment of affine symmetric spaces. Applying the results of Chapter II, we shall first clarify the existence and properties of a natural affine connection on any symmetric homogeneous space. We include a generalization of Cartan's theorem concerning the largest connected group of isometries of a riemannian symmetric space (Theorem 16. 1). We shall then give a theorem showing the equivalence of various definitions of affine symmetric spaces (Theorem 17. 1); the proof of the most essential part is given as a special case of a more general theorem which is established in Chapter IV. In Chapter IV, we shall characterize the spaces with affine connection such that the covariant derivatives of the torsion and curvature tensor fields are both zero (Theorem 18. 1). It turns out that such a space may be characterized in two ways, geometrically and group-theoretically. Geometrically, it is an affine connection which is invariant by its parallelism. Group-theoretically, such a space may be locally representable as a reductive homogeneous space with its canonical affine connection of the second kind.

My interest in the subject arose from Dr. Koszul's seminar on symmetric spaces given at the University of Chicago in the spring of 1952, and from the subsequent conversations I had with Dr. Koszul and Professor Chern. I have also had the privilege of having frequent discussions and receiving valuable criticism from Professor Chevalley during the preparation of this work. I would like to take this opportunity to express my sincere gratitude to these mathematicians.

Chapter I. Affine Connection.

1. Affine connection. Let M be a connected differentiable manifold of class C^∞. We denote by $\mathfrak{F}$ the ring of all differentiable functions on M and

by $\mathfrak{X}$ the set of all infinitesimal transformations on M.[1] Xf, where $X \in \mathfrak{X}$ and $f \in \mathfrak{F}$, denotes the function in $\mathfrak{F}$ which is defined by $(Xf)(p) = X_p(f)$. On the other hand, fX denotes the infinitesimal transformation which is defined by $(fX)_p = f(p)X_p$. In this way, $\mathfrak{X}$ can be considered as a module over $\mathfrak{F}$. $\mathfrak{X}$ can be also regarded as a Lie algebra of infinite dimension over the reals with the usual bracket operation $[X, Y]$. We then have

$$[fX, gY] = fg[X, Y] + f(Xg)Y - g(Yf)X \tag{1.1}$$

for $f, g \in F$ and $X, Y \in \mathfrak{X}$.

An *affine connection* on M is, by definition,[2] a rule which assigns to each $X \in \mathfrak{X}$ an $\mathfrak{F}$-endomorphism $t(X)$ of $\mathfrak{X}$ and which satisfies the following conditions:

$$\text{(a)} \quad t(X_1 + X_2) = t(X_1) + t(X_2) \tag{1.2}$$

$$\text{(b)} \quad t(fX)(Y) = ft(X)(Y) + (Yf)X.$$

Then the following condition is automatically satisfied.

(c) If either X or Y is identically zero in an open subset U of M, then so is $t(X)(Y)$.

If M has an affine connection, then any open submanifold U of M has an induced affine connection. Namely, one can define an affine connection t_U in U as follows: Let X and Y be any given infinitesimal transformations in U. For any given point p in U, there exist infinitesimal transformations X' and Y' on M which coincide in a certain neighborhood of p with X and Y respectively. One defines $(t_U(X)(Y))_p = (t(X')(Y'))_p$, which is a tangent vector at p. This tangent vector is uniquely determined independently of the choice of extensions X' and Y' of X and Y, in virtue of the condition (c). It is a trivial matter to verify that $p \to (t_U(X)(Y))_p$ is an infinitesimal transformation in U and that $X \to t_U(X)$ satisfies the conditions of affine connection in U.

In particular, if U is a coordinate neighborhood with local coordinates $x^1, x^2, \cdots, x^n$, the induced affine connection t_U in U can be expressed by

$$t_U(\partial/\partial x^i)(\partial/\partial x^j) = \sum_{k=1}^{n} \Gamma^k_{ji}(\partial/\partial x^k), \tag{1.3}$$

---

[1] For the terminology concerning manifolds we follow Chapter III of [7] with obvious change of analyticity into infinite differentiability.

[2] This definition is due to Koszul and has been communicated to me orally. For the standard treatment of affine connections, we refer to [6].

where $(\partial/\partial x^i)$ are infinitesimal transformations in U such that $(\partial/\partial x^i)f = \partial f/\partial x^i$ for any differentiable function f in U and $\Gamma^k_{ji}(i, j, k = 1, 2, \cdots, n)$ are certain differentiable functions in U. By a coordinate transformation in U the Γ^k_{ji} undergo the transformation which is well known in classical differential geometry; this follows easily from the conditions of affine connection. Conversely, given an affine connection in the classical sense by means of Γ^k_{ji} in each coordinate neighborhood, one defines t_U by (1. 3) in each coordinate neighborhood and then defines $(t(X)(Y))_p = (t_U(X_U)(Y_U))_p$ at each point p of M, where U is a coordinate neighborhood of p and X_U and Y_U are the restriction of X and Y to U respectively. It is easy to see that this process defines legitimately an affine connection $X \to t(X)$ in our sense. We have thus seen that our notion of affine connection coincides with the classical one.

2. Transformations. Let ϕ be a differentiable homeomorphism of M onto itself. We shall denote by $(\phi')_p$ the linear mapping of the tangent space $T(p)$ at p into the tangent space $T(\phi(p))$ at $\phi(p)$ which is induced by ϕ. We shall also denote by ϕ' the mapping of $\mathfrak{X}$ onto itself defined by

$$(\phi' \cdot X)_p = (\phi')_{\phi^{-1} \cdot p} \cdot X_{\phi^{-1} \cdot p},$$

as well as the mapping of $\mathfrak{F}$ onto itself defined by

$$(\phi' \cdot f)(p) = f(\phi^{-1}(p)).$$

Then we have the following formulas:

$$\phi' \cdot (Xf) = (\phi' \cdot X)(\phi' \cdot f) \tag{2. 1}$$

$$\phi' \cdot (fX) = (\phi' \cdot f)(\phi' \cdot X) \tag{2. 2}$$

for $f \in \mathfrak{F}$ and $X \in \mathfrak{X}$. X is called invariant by ϕ if $\phi' \cdot X = X$.

Given an affine connection $X \to t(X)$ on M and a differentiable homeomorphism ϕ of M onto itself, one can define a new affine connection $X \to t'(X)$ by

$$t'(X)(Y) = (\phi^{-1})' \cdot (t(\phi' \cdot X)(\phi' \cdot Y)). \tag{2. 3}$$

If $t(X) = t'(X)$ for every $X \in X$, we say that the affine connection is *invariant* by ϕ, or that ϕ is an *affine transformation* of M (with respect to the given affine connection).

Let us consider the case where the $\mathfrak{F}$-module $\mathfrak{X}$ has a finite base $X_1, X_2, \cdots, X_n$, that is, every $X \in \mathfrak{X}$ can be uniquely expressed as $X = \sum_{i=1}^{n} f_i X_i$, where f_i $(i = 1, 2, \cdots, n)$ are functions in $\mathfrak{F}$. Then an affine connection

$X \to t(X)$ is uniquely determined by the values (in $\mathfrak{X}$) of $t(X_i)(X_j)$ $(i, j = 1, 2, \cdots, n)$ as we see from the conditions of affine connection. Conversely, by assigning any arbitrary values (in $\mathfrak{X}$) to $t(X_i)(X_j)$ $(i, j = 1, 2, \cdots, n)$ one can define $t(X)(Y)$ for all X and Y in $\mathfrak{X}$ so that $X \to t(X)$ is an affine connection. Let furthermore ϕ be a differentiable homeomorphism of M onto itself. Then an affine connection $X \to t(X)$ is invariant by ϕ if and only if $\phi' \cdot (t(X_i)(X_j)) = t(\phi' \cdot X_i)(\phi' \cdot X_j)$ for all $i, j = 1, 2, \cdots, n$. In particular, if each X_i is invariant by ϕ, then the affine connection which is determined by $t(X_i)(X_j)$ is invariant by ϕ if and only if all $t(X_i)(X_j)$ are invariant by ϕ.

In order to illustrate an application of the above considerations, we shall see how the left invariant affine connections on a connected Lie group are determined. Let G be a connected Lie group of dimension n and let $X_1, X_2, \cdots, X_n$ be a set of linearly independent left invariant infinitesimal transformations of G. They form an $\mathfrak{F}$-base of the space $\mathfrak{X}$ over G. A left invariant affine connection on G means an affine connection which is invariant by the left translations of G. By the above remarks, any left invariant affine connection is determined by its values of $t(X_i)(X_j)$ $(i, j = 1, 2, \cdots, n)$ in $\mathfrak{X}$ which are themselves left invariant infinitesimal transformations. Since the left invariant infinitesimal transformations are determined by their values at the identity element e, that is, by the corresponding elements of the Lie algebra $\mathfrak{g}$ of G, we see that each left invariant affine connection gives rise to a bilinear function α on $\mathfrak{g} \times \mathfrak{g}$ with values in $\mathfrak{g}$ and vice versa.

(2.4) *There is a one-to-one correspondence between the set of left invariant affine connections on G and the set of bilinear functions α on $\mathfrak{g} \times \mathfrak{g}$ with values in $\mathfrak{g}$, the correspondence being given by $(t(X)(Y))_e = \alpha(Y, X)$ for $X, Y \in \mathfrak{g}$.*

3. Torsion and curvature. We shall find the expressions for the torsion and curvature tensor fields in terms of our definition of affine connection. Let U be a coordinate neighborhood in M with local coordinates $x^1, x^2, \cdots, x^n$. Let $X_1, X_2, \cdots, X_n$ be any base of the module $\mathfrak{X}(U)$ of infinitesimal transformations in U such that $\partial/\partial x^i = \sum_{p=1}^{n} u_{ip} X_p$ $(i = 1, 2, \cdots, n)$ where u_{ip} are certain differentiable functions in U with the determinant $|u_{ip}|$ which is not zero at any point of U. If we set $t(\partial/\partial x^i)(\partial/\partial x^j) = \sum_{k=1}^{n} \Gamma^k_{ji}(\partial/\partial x^k)$ as before, the torsion tensor field has coefficients $T^k_{ij} = \Gamma^k_{ij} - \Gamma^k_{ji}$. We have easily

$$(3.1)\qquad \sum_{k=1}^{n} T^k{}_{ij}(\partial/\partial x^k) = t(\partial/\partial x^j)(\partial/\partial x^i) - t(\partial/\partial x^i)(\partial/\partial x^j)$$
$$= \sum_{p,q=1}^{n} u_{jq}u_{ip}\{t(X_q)(X_p) - t(X_p)(X_q) - [X_p, X_q]\}.$$

This leads us to the definition of the following mapping T of $\mathfrak{X} \times \mathfrak{X}$ into $\mathfrak{X}$:

$$(3.2)\qquad T(X, Y) = t(Y)(X) - t(X)(Y) - [X, Y], \qquad X, Y \in \mathfrak{X}.$$

It has the properties that $T(X, Y) = -T(Y, X)$ and $T(fX, gY) = fgT(X, Y)$ where $f, g \in \mathfrak{F}$. It is a tensor field of type $(1, 2)$. Since

$$T(\partial/\partial x^i, \partial/\partial x^j) = \sum_{k=1}^{n} T^k{}_{ij}(\partial/\partial x^k),$$

we see that T is nothing but the torsion tensor field.

As for the curvature tensor field, we have the coefficients

$$R^i{}_{ijk} = \partial\Gamma^i{}_{kl}/\partial x^j - \partial\Gamma^i{}_{jl}/\partial x^k + \sum_{m=1}^{n} (\Gamma^i{}_{jm}\Gamma^m{}_{kl} - \Gamma^i{}_{km}\Gamma^m{}_{jl})$$

which may be expressed in the following fashion:

$$(3.3)\qquad \sum_{i=1}^{n} R^i{}_{ijk}(\partial/\partial x^i) = t(t(\partial/\partial x^l)(\partial/\partial x^k))(\partial/\partial x^j)$$
$$- t(t(\partial/\partial x^l)(\partial/\partial x^j))(\partial/\partial x^k).$$

We define therefore a mapping R of $\mathfrak{X} \times \mathfrak{X} \times \mathfrak{X}$ into $\mathfrak{X}$, or what amounts to the same thing, a mapping of $\mathfrak{X} \times \mathfrak{X}$ into the space of mappings of $\mathfrak{X}$ into itself by

$$(3.4)\qquad R(X, Y) \cdot Z = t(t(Z)(Y))X - t(t(Z)(X))Y - t(Z) \cdot [X, Y]$$

for X, Y, Z in $\mathfrak{X}$. R has the properties that $R(X, Y) = -R(Y, X)$ and $R(fX, gY) \cdot hZ = fgh\, R(X, Y) \cdot Z$ for X, Y, Z in $\mathfrak{X}$ and f, g, h in $\mathfrak{F}$. It is a tensor field of type $(1. 3)$. We see from $(3. 3)$ that R is nothing but the curvature tensor field.

4. Covariant differentiation. The covariant differentiation Δ which arises from a given affine connection is the derivation of the algebra of tensor fields on M which is uniquely determined by the following conditions:

$(4. 1)$ (a) $\Delta f = df$ *for* $f \in \mathfrak{F}$, *where* d *is the exterior differentiation*;

(b) *For any* $X \in \mathfrak{X}$ *and for any differential form* ω *of degree* 1, $\langle \Delta X, \omega\rangle + \langle X, \Delta\omega\rangle = d\langle X, \omega\rangle$, *where* $\langle\ ,\ \rangle$ *denotes the contraction of tensors*;

(c) *ΔX for $X \in \mathfrak{X}$ is determined as follows: in any coordinate neighborhood U with coordinates $x^1, x^2, \cdots, x^n$*

$$\Delta(\partial/\partial x^i) = \sum_{j,k=1}^{n} \Gamma^k{}_{ji}(\partial/\partial x^k) \otimes dx^j,$$

where $\otimes$ denotes the tensor product.

From the last condition (c) and (1.3) we see that $\Delta(\partial/\partial x^i) = t(\partial/\partial x^i)$ and hence

$$\Delta(X) = t(X) \text{ for } X \in \mathfrak{X}. \tag{4.2}$$

For each $Y \in \mathfrak{X}$ we introduce a mapping Δ_Y of $\mathfrak{X}$ into itself by

$$\Delta_Y \cdot X = t(X)Y, \qquad X \in \mathfrak{X}. \tag{4.3}$$

It is easy to see that the value $\Delta_Y \cdot X$ at a point p depends, for each X, only on the value of Y at p. (3.4) may be written as

$$R(X, Y) = [\Delta_X, \Delta_Y] - \Delta_{[X,Y]}. \tag{4.4}$$

This shows that the linear mapping $X \to \Delta_X$ of the Lie algebra $\mathfrak{X}$ into the Lie algebra of endomorphisms of the linear space $\mathfrak{X}$ is not in general a homomorphism of Lie algebras and the 'defect' is given by the curvature tensor field.

Without difficulty, we can find the expression for ΔR as follows:

$$\begin{aligned}\Delta R(X, Y; Z) \cdot W = {} & \Delta_Z \cdot (R(X, Y) \cdot W) - R(X, Y) \cdot (\Delta_Z \cdot W) \\ & - R(\Delta_Z \cdot X, Y) \cdot W - R(X, \Delta_Z \cdot Y) \cdot W.\end{aligned} \tag{4.5}$$

We write this expression also as $((\Delta_Z \cdot R)(X, Y)) \cdot W$. This corresponds to the classical formula:

$$\begin{aligned}R^i{}_{ljk,s} = {} & \partial R^i{}_{ljk}/\partial x^s + \sum_{p=1}^{n} \Gamma^i{}_{sp} R^p{}_{ljk} \\ & - \sum_{p=1}^{n} \Gamma^p{}_{sl} R^i{}_{pjk} - \sum_{p=1}^{n} \Gamma^p{}_{sj} R^i{}_{lpk} - \sum_{p=1}^{n} \Gamma^p{}_{sk} R^i{}_{ljp}.\end{aligned}$$

5. Parallelism. By a *regular curve* in M we shall mean a continuous mapping $x(s)$ of the unit interval $[0, 1]$ into M which can be extended to a differentiable mapping of an open interval $(-\epsilon, 1 + \epsilon)$ for some $\epsilon > 0$ and whose image is a (1-dimensional) submanifold of M in the sense of [7]. By a *curve* we shall mean a continuous mapping of the unit interval into M which is composed of a finite number of regular curves. The parallelism is first defined along regular curves lying in a coordinate neighborhood and then,

by obvious extension, along any curves in M. Let $x(s)$ be a regular curve lying in a coordinate neighborhood U, and let X_s be the family of tangent vectors to the curve. A given family Y_s of tangent vectors at the points $x(s)$ is said to be *parallel* along $x(s)$ if

$$t(Y)(X) = 0 \text{ on } x(s), \tag{5.1}$$

where X and Y are infinitesimal transformations defined in U which coincide with X_s and Y_s respectively on $x(s)$. Such extensions X and Y always exist and the condition (5.1) does not depend on the choice of extensions. It is easy to show that this notion of parallelism coincides with the classical one.

A *path* is defined to be a regular curve such that the family of tangent vectors X_s to the curve is parallel along the curve itself. For any point p of M and for any given tangent vector X_p at p, there exists a path $x(s)$, defined for $-\epsilon < s < \epsilon$ for some $\epsilon > 0$, such that $x(0) = p$ and such that the tangent vector at p to $x(s)$ is equal to the given X_p, that is, $\lim_{s \to 0} (1/s)(f(x(s)) - f(p)) = X_p f$ for any function f. Such a parameter s is uniquely determined and is called the canonical parameter of the path $x(s)$. M is called *complete* if every path can be extended for any large value of its canonical parameter.

6. Holonomy groups. We shall not go into the concept of development of frames along a curve, which enables us to define the holonomy group associated to the given affine connection. When we choose an arbitrary point p, the holonomy group is realized as the group of affine transformations of the affine tangent space at p, that is, the tangent space $T(p)$ provided with a natural affine structure. However, the homogeneous part of the holonomy group can be defined only with the help of parallelism. In the present paper we shall agree that the word *holonomy group* means the homogeneous part of the whole holonomy group. Thus the holonomy group H^* is the linear group in the tangent space consisting of the parallel displacements corresponding to all closed curves. The subgroup H^*_0 of H^* which corresponds to all closed curves homotopic to 0 will be called the *restricted holonomy group* of M.

If M is n-dimensional, the holonomy group H^* and the restricted holonomy group H^*_0 are subgroups of $GL(n, R)$. It has been proved recently by several authors that H^*_0 is the identity component of H^* and that H^*_0 is a connected Lie group. We shall state a theorem of Chevalley which shows how the Lie algebra of the restricted hononomy group is deter-

mined by the curvature tensor field and all its successive covariant differentials.

For each point p of M, let $\mathfrak{h}^*(p)$ be the Lie algebra of endomorphisms of the tangent space at p which is generated by all endomorphisms of the forms $(R(X, Y))_p$, $((\Delta_Z \cdot R)(X, Y))_p$, $((\Delta_W \cdot \Delta_Z \cdot R)(X, Y))_p, \cdots$, where X, Y, Z, $W, \cdots$ denote arbitrary infinitesimal transformations. Let $r(p) = \dim \mathfrak{h}^*(p)$ and $r(M) = \max_{p \in M} r(p)$. If $r(p) = r(M)$, p is called a regular point. Then, *if the set of regular points of M is connected, the restricted holonomy group has the Lie algebra which is isomorphic with $\mathfrak{h}^*(p)$ for any regular point p* (Chevalley).[3]

If M admits a transitive group of affine transformations, then all $\mathfrak{h}^*(p)$, $p \in M$, are isomorphic with each other so that M consists of regular points. Therefore, the restricted holonomy group has the Lie algebra isomorphic with $\mathfrak{h}^*(p)$ for any point $p \in M$. We shall use this result in Chapter II.

Chapter II. Invariant Affine Connections on Homogeneous Spaces.

7. Reductive homogeneous spaces. A homogeneous space G/H of a connected Lie group G is called *reductive* if the following condition is satisfied: in the Lie algebra $\mathfrak{g}$ of G there exists a subspace $\mathfrak{m}$ such that $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$ (direct sum of vector subspaces) and $ad(h)\mathfrak{m} \subset \mathfrak{m}$ for all $h \in H$, where $\mathfrak{h}$ is the subalgebra of $\mathfrak{g}$ corresponding to the identity component H_0 of H and $ad(h)$ denotes the adjoint representation of H in $\mathfrak{g}$. G/H is reductive in either of the following cases:

(a) H is compact;

(b) H is connected and the subalgebra $\mathfrak{h}$ is reductive in $\mathfrak{g}$, that is, $ad(\mathfrak{h})$ in $\mathfrak{g}$ is completely reducible. This is the case if H is connected and semi-simple.

(c) H is a discrete subgroup of G; $\mathfrak{h} = 0$ and $\mathfrak{m} = \mathfrak{g}$.

From now on, we shall always assume a homogeneous space to be reductive and shall consider a fixed decomposition $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$ of the Lie algebra satisfying the above condition.

G/H is considered to be the space of cosets xH, $x \in G$, on which G acts to the left, that is, $a \in G$ maps the point xH into the point axH. We denote by p_0 the point represented by the coset H. Then the subspace $\mathfrak{m}$ can be identified with the tangent space at p_0. Let $X_1, X_2, \cdots, X_n$ be a base of $\mathfrak{g}$ such that the first m elements span $\mathfrak{m}$ and the last $n - m$ elements span $\mathfrak{h}$.

[3] This result is unpublished.

As is well known,[4] there exists a neighborhood V of the identity element e of G with a canonical system of coordinates $x^1, x^2, \cdots, x^n$ such that V is the topological product of the sets

$$N = \{x \in V \mid x^{m+1} = x^{m+2} = \cdots x^n = 0\}$$

and

$$K = \{x \in V \mid x^1 = x^2 = \cdots x^m = 0\},$$

where K is contained in H_0. If we denote by π the canonical projection of G onto G/H, π is a differentiable homeomorphism of N onto a neighborhood N^* of p_0 in G/H. We have $\pi(c) = \tau(c) \cdot p_0$ for $c \in N$, where $\tau(c)$ denotes the transformation of G/H which is induced by c. N^* is the set of points of the form $\tau(c) \cdot p_0$, $c \in N$.

For each element X of $\mathfrak{m}$, we define an infinitesimal transformation X^* in N^* as follows:

$$(7.1) \qquad (X^*)_{\tau(c)\cdot p_0} = (\tau(c))'_{p_0} \cdot X \qquad (c \in N),$$

where X in the right hand side denotes the tangent vector at p_0 which is identified with the given $X \in \mathfrak{m}$. $(X^*)_{p_0}$ is nothing but this element X of the tangent space at p_0. Since $\tau(c)$ is a differentiable homeomorphism of G/H onto itself and since N and N^* are differentiably homeomorphic by π, we see that X^* so defined is an infinitesimal transformation in N^*. Since $X_1, X_2, \cdots, X_m$ is a base of $\mathfrak{m}$, the corresponding $X_1^*, X_2^*, \cdots, X_m^*$ are linearly independent at every point of N^*, that is, they form a base of the module of infinitesimal transformations on N^*.

Let h be any element of H. Since $\tau(h) \cdot p_0 = p_0$, $\tau(h)$ induces a linear transformation $(\tau(h))'_{p_0}$ of the tangent space at p_0 onto itself, which is the same as $ad(h)$ on $\mathfrak{m}$. We shall show how each infinitesimal transformation X^* $(X \in \mathfrak{m})$ is transformed by $\tau(h)$. Take a subset N_1 of N such that $hN_1h^{-1} \subset N$ (the existence of such N_1 follows from the condition $ad(h)\mathfrak{m} \subset \mathfrak{m}$) and let $N_1^* = \{\tau(c) \cdot p_0 \mid c \in N_1\}$. Then we have $\tau(h) \cdot N_1^* \subset N^*$ and

$$(7.2) \qquad (\tau(h))' \cdot X^* = (ad(h)X)^* \text{ on } \tau(h) \cdot N_1^*.$$

In fact, we have for any $c \in N_1$

$$\begin{aligned}((\tau(h))' \cdot X^*)_{\tau(h)\tau(c)\cdot p_0} &= (\tau(h))' \cdot (X^*)_{\tau(c)\cdot p_0} = (\tau(hc))' \cdot X \\ &= (\tau(hch^{-1}))' \cdot (\tau(h))' \cdot X = (\tau(hch^{-1}))' \cdot (ad(h)) \cdot X \\ &= (ad(h)X)^*_{\tau(h)\tau(c)\cdot p_0},\end{aligned}$$

since $hch^{-1} \in N$ and $\tau(hch^{-1}) \cdot p_0 = \tau(h) \cdot \tau(c) \cdot p_0$.

For later use, we shall prove

---

[4] For example, see Chapter IV, § V of [7].

(7.3) *For any X and Y in $\mathfrak{m}$, $[X^*, Y^*]_{p_0} = [X, Y]_{\mathfrak{m}}$, where $[X, Y]_{\mathfrak{m}}$ denotes the $\mathfrak{m}$-component of the element $[X, Y] \in \mathfrak{g} = \mathfrak{m} + \mathfrak{h}$.*

In fact, let $x(t)$ and $y(s)$ be the 1-parameter subgroup of G generated by X and Y respectively. For any function f in N^*, we have

$$(Y^*f)_{x^*(t)} = \lim_{s \to 0} (1/s)\{(f \circ \pi)(x(t)y(s)) - (f \circ \pi)(x(t))\},$$

where $x^*(t)$ denotes the image of $x(t)$ by π, and

$$X^*_{p_0}(Y^*f) = \lim_{t \to 0} (1/t)\{(Y^*f)(x^*(t)) - (Y^*f)(p_0))\}.$$

From this expression and a similar one for $Y^*_{p_0}(X^*f)$, we see that $[X^*, Y^*]_{p_0}f = [X, Y]_e(f \circ \pi) = [X, Y]_{\mathfrak{m}}f$, which proves (7.3).

8. Fundamental existence theorem. Keeping the notation of the preceding section, we shall now prove the fundamental existence theorem for invariant affine connections on G/H. An invariant affine connection means an affine connection on G/H which is invariant by all $\tau(a)$, $a \in G$.

THEOREM 8.1. *Let G/H be a reductive homogeneous space with a fixed decomposition of the Lie algebra $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$, $ad(H)\mathfrak{m} \subset \mathfrak{m}$. There exists a one-to-one correspondence between the set of all invariant affine connections on G/H and the set of all bilinear functions α on $\mathfrak{m} \times \mathfrak{m}$ with values in $\mathfrak{m}$ which are invariant by $ad(H)$, that is, $ad(h) \cdot \alpha(X, Y) = \alpha(ad(h)X, ad(h)Y)$ for $X, Y \in \mathfrak{m}$ and $h \in H$. The correspondence is given by*

$$\alpha(X, Y) = (t(Y^*)(X^*))_{p_0}. \tag{8.1}$$

For the sake of simplicity, we shall call such a bilinear function α on $\mathfrak{m} \times \mathfrak{m}$ a *connection function* on $\mathfrak{m} \times \mathfrak{m}$ (or, more precisely, a connection function with respect to the decomposition $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$).

Proof. Assume first that there exists an invariant affine connection t on G/H. We define a bilinear function $\alpha(X, Y)$ on $\mathfrak{m} \times \mathfrak{m}$ with values in $\mathfrak{m}$ by (8.1). Let h be any fixed element of H, and take a subset N_1 of N such that (7.2) holds. Since the affine connection is invariant by G, the induced affine connection in N_1 is invariant by $\tau(h)$, that is, we have by (2.3)

$$(\tau(h))' \cdot (t(Y^*)(X^*)) = t((\tau(h))' \cdot Y^*)((\tau(h))' \cdot X^*)$$

on $\tau(h) \cdot N_1{}^*$. Taking the values of both sides at p_0 and using **(7.2)**, we obtain

$$(\tau(h))' \cdot (t(Y^*)(X^*))_{p_0} = (\tau(h))' \cdot \alpha(X, Y) = ad(h) \cdot \alpha(X, Y)$$

and

$$(t((\tau(h))'\cdot Y^*)((\tau(h))'\cdot X^*))_{p_0} = (t((ad(h)Y)^*)(ad(h)X)^*))_{p_0}$$
$$= \alpha(ad(h)X, ad(h)Y),$$

which prove the invariance of $\alpha(X, Y)$ by $ad(h)$.

Conversely, we shall prove that any connection function α on $\mathfrak{m}\times\mathfrak{m}$ gives rise to an invariant affine connection which is related to the original α by (8. 1). For a base $X_1, X_2, \cdots, X_m$ of $\mathfrak{m}$, the corresponding $X_1^*, X_2^*, \cdots, X_m^*$ form a base of the module of infinitesimal transformations in N^*. As we have remarked in Section 2, there exists an affine connection t^* in N^* such that

$$(t^*(X_j^*)(X_i^*))_{\tau(c)\cdot p_0} = (\alpha(X_i, X_j))^*_{\tau(c)\cdot p_0} \qquad (c\in N).$$

Since α is invariant by $ad(H)$, it follows from (7. 2) that

(8. 2) $\quad (\tau(h))'\cdot(t^*(X_j^*)(X_i^*))_{p_0} = (t^*((\tau(h))'\cdot X_j^*)((\tau(h))'\cdot X_i^*))_{p_0}$

for any $h\in H$. More generally, we have

(8. 3) *For any infinitesimal transformations $X^\dagger$ and $Y^\dagger$ in N^*,*

$$(\tau(h))'\cdot(t^*(Y^\dagger)(X^\dagger))_{p_0} = (t^*((\tau(h))'\cdot Y^\dagger)((\tau(h))'\cdot X^\dagger))_{p_0}.$$

This can be proved easily by using (1. 2), (2. 1), (2. 2) and (8. 2). Now we define an affine connection t in G/H as follows.

For any infinitesimal transformation $X^\dagger$ and $Y^\dagger$ on G/H, we define

$$(t(Y^\dagger)(X^\dagger))_p = (\tau(a))'\cdot(t^*((\tau(a^{-1}))'\cdot Y^\dagger))((\tau(a^{-1}))'\cdot X^\dagger))_{p_0}$$

at any point $p = \tau(a)\cdot p_0$ $(a\in G)$ of G/H. (8. 3) implies that this tangent vector at p is independent of the choice of an element $a\in G$ such that $p = \tau(a)\cdot p_0$. It is easy to show that $p\to(t(Y^\dagger)(X^\dagger))_p$ is an infinitesimal transforamtion on G/H and that $Y^\dagger\to t(Y^\dagger)$ defines an affine connection on G/H. For example, we have for any differentiable function f on G/H (setting $p = \tau(a)\cdot p_0$ and $\phi = \tau(a^{-1})$)

$$\begin{aligned}(t(Y^\dagger)(fX^\dagger))_p &= (\phi^{-1})'\cdot(t^*(\phi'\cdot Y^\dagger)(\phi'\cdot(fX^\dagger))_{p_0}\\ &= (\phi^{-1})'\cdot(t^*(\phi'\cdot Y))((\phi'\cdot f)(\phi'\cdot Y^\dagger))_{p_0}\\ &= (\phi^{-1})'\cdot((\phi'\cdot f)(t^*(\phi'\cdot Y^\dagger))(\phi'\cdot X^\dagger))_{p_0}\\ &= f(p)(\phi^{-1})'\cdot(t^*(\phi'\cdot Y^\dagger)(\phi'\cdot X^\dagger))_{p_0} = (ft(Y^\dagger)(X^\dagger))_p,\end{aligned}$$

that is, $t(Y^\dagger)(fX^\dagger) = ft(Y^\dagger)(X^\dagger)$. The other conditions of affine connection can be proved similarly. Finally, the invariance of this affine connection t is clear from the definition. Since

$$(t(Y^*)(X^*))_{p_0} = (t^*(Y^*)(X^*))_{p_0} = \alpha(X, Y),$$

t is related to the given α by (8. 1).

It is also clear that an invariant affine connection on G/H is uniquely determined by the corresponding connection function α. We have thereby completed the proof.

9. Torsion and curvature of invariant affine connections. Let t be the invariant affine connection on G/H which is determined by a connection function α on $\mathfrak{m} \times \mathfrak{m}$. We shall find the expressions of the torsion and curvature tensor fields of t in terms of α. We remark that these tensor fields are invariant by G and hence are determined by their values at p_0. The covariant differentiation commutes with $\tau(a)$, $a \in G$, that is, $\tau(a) \cdot \Delta \cdot F = \Delta \cdot \tau(a) \cdot F$ for every $a \in G$ and for every tensor field F.

The torsion tensor field T is easily determined. Let X^* and Y^* be the infinitesimal transformations which correspond to X and Y in $\mathfrak{m}$ respectively. We have from (3. 2) that $T(X^*, Y^*) = t(Y^*)(X^*) - t(X^*)(Y^*) - [X^*, Y^*]$. It value at p_0 is then equal to $\alpha(X, Y) - \alpha(Y, X) - [X, Y]_{\mathfrak{m}}$ in virtue of (7. 3) and (8. 1). We shall denote this expression by

$$T(X, Y) = \alpha(X, Y) - \alpha(Y, X) - [X, Y]_{\mathfrak{m}} \qquad (X, Y \in \mathfrak{m}). \tag{9. 1}$$

It is a bilinear function on $\mathfrak{m} \times \mathfrak{m}$ with values in $\mathfrak{m}$.

As for the curvature tensor field R, we have by (3. 4)

$$\begin{aligned} &R(X^*, Y^*) \cdot Z^* \\ &\quad = t(t(Z^*)(Y^*))X^* - t(t(Z^*)(X^*))Y^* - t(Z^*) \cdot [X^*, Y^*]. \end{aligned} \tag{9. 2}$$

In order to calculate this, let $X_1, X_2, \cdots, X_m$ be a base of $\mathfrak{m}$. We set for any $p = \tau(c) \cdot p_0$ $(c \in N)$

$$((\tau(c^{-1})' \cdot X_i^*)_q = \sum_{\mu=1}^{m} \phi_{\mu i}(p, q) X_\mu^*, \ q \in N^*, \qquad i = 1, 2, \cdots, m,$$

where $\phi_{\mu i}$ are differentiable functions in p and q which satisfy the following condition:

$$\phi_{\mu i}(p, p_0) = \phi_{\mu i}(p_0, q) = 0 \text{ for any } p, q \text{ in } N^*. \tag{9. 3}$$

Then we have

$$\begin{aligned} (t(X_l^*)(X_k^*))_p &= (\tau(c))' \cdot (t^*((\tau(c^{-1}))' \cdot X_l^*)((\tau(c^{-1}))' \cdot X_k^*))_{p_0} \\ &= (\tau(c))' \cdot (t^*(\sum_{\nu=1}^{m} \phi_{\nu l} X_\nu^*)(\sum_{\mu=1}^{m} \phi_{\mu k} X_\mu^*))_{p_0} \\ &= (\tau(c))' \cdot \{\sum_{\mu,\nu=1}^{m} \phi_{\nu l}(p, p_0)\phi_{\mu k}(p, p_0)\alpha(X_\mu, X_\nu) \\ &\qquad + \sum_{\mu,\nu=1}^{m} \phi_{\mu k}(p, p_0)(X_\mu^* \phi_{\nu l}(p, p_0))X_\nu\} \\ &= (\alpha(X_k, X_l))_p^* + \sum_{\nu=1}^{m} (X_k^* \phi_{\nu l})(p, p_0)(X_\nu^*)_p, \end{aligned}$$

where $(X_k^*\phi_{\nu l})(p, p_0)$ is the function of p which is the result of applying X_k^* to the second variable of $\phi_{\nu l}(p, q)$ at $q = p_0$. We have, by a similar computation as above

$$\begin{aligned}(t(t(X_l^*)(X_k^*))X_j^*)_{p_0} &= \{t(\alpha(X_k, X_l)^* + \sum_{\nu=1}^{m}(X_k^*\phi_{\nu l})(p, p_0)X_\nu^*)X_j^*\}_{p_0} \\ &= \alpha(X_j, \alpha(X_k, X_l)) + \sum_{\nu=1}^{m}(X_k^*\phi_{\nu l})(p_0, p_0)\alpha(X_j, X_k) \\ &\quad + \sum_{\nu=1}^{m}(X_j^*)_{p_0}(X_k^*\phi_{\nu l})(p, p_0)X_\nu \\ &= \alpha(X_j, \alpha(X_k, X_l)) + \sum_{\nu=1}^{m}(X_j^*)_{p_0}(X_k^*\phi_{\nu l})(p, p_0)\cdot X_\nu,\end{aligned}$$

since $(X_k^*\phi_{\nu l})(p, p_0) = 0$ by (9. 3). In the last expression $(X_j^*)_{p_0}(X_k^*\phi_{\nu l})(p, p_0)$ denotes the result of applying X_j^* at $p = p_0$ to the function $(X_j^*\phi_{\nu l})(p, p_0)$. Hence we get

$$\begin{aligned}(9.4)\quad &\{(t(t(X_l^*)(X_k^*))X_j^*) - (t(t(X_l^*)(X_j^*))X_k^*)\}_{p_0} \\ &= \alpha(X_j, \alpha(X_k, X_l)) - \alpha(X_k, \alpha(X_j, X_l)) \\ &\quad + \sum_{\nu=1}^{m}\{(X_j^*)_{p_0}(X_k^*\phi_{\nu l})(p, p_0) - (X_k^*)_{p_0}(X_j^*\phi_{\nu l})(p, p_0)\}X_\nu.\end{aligned}$$

We shall show that the last term is equal to $-[[X_j, X_k]_{\mathfrak{h}}, X_l]$, where $[X_j, X_k]_{\mathfrak{h}}$ denotes the $\mathfrak{h}$-component of the element $[X_j, X_k] \in \mathfrak{g}$. Let $p = \tau(c)\cdot p_0$ and $q = \tau(c')\cdot p_0$ with $c, c' \in N$. Then

$$\begin{aligned}((\tau(c^{-1}))'\cdot X_l^*)_q &= (\tau(c^{-1}))'\cdot(X_l^*)_{\tau(cc')\cdot p_0} \\ &= (\tau(c^{-1}))'\cdot(\tau(c''))'\cdot X_l = (\tau(c'))'\cdot(\tau(h^{-1}))'\cdot X_l \\ &= (\tau(c'))'\cdot(ad(h^{-1})X_l),\end{aligned}$$

where $c'' \in N$ and $h \in K$ are uniquely determined elements such that $cc' = c''h$ in the decomposition of V into the product of N and K explained in Section 7. Therefore we see that $(ad(h^{-1})X_l)^*_q = \sum_{\nu=1}^{m}\phi_{\nu l}(p, q)(X_\nu^*)_q$. It is not difficult to derive from this that

$$(9.5)\qquad \sum_{\nu=1}^{m}\{(X_j^*)_{p_0}(X_k^*\phi_{\nu l})(p, p_0) - (X_k^*)_{p_0}(X_j^*\phi_{\nu l})(p, p_0)\}X_\nu = -[[X_j, X_k]_{\mathfrak{h}}, X_l].$$

Finally the term $(t(X_l^*)\cdot[X_j^*, X_k^*])_{p_0}$ is easily seen to be equal to $\alpha([X_j, X_k]_{\mathfrak{m}}, X_l)$. Using this and (9. 4), (9. 5) and (9. 2), we obtain the following expression for the value at p_0 of the curvature tensor field

$$(9.6)\quad R(X,Y)\cdot Z = \alpha(X,\alpha(Y,Z)) - \alpha(Y,\alpha(X,Z)) - \alpha([X,Y]_{\mathfrak{m}},Z) - [[X,Y]_{\mathfrak{h}},Z], \qquad X,Y,Z \in \mathfrak{m}.$$

The covariant differential ΔR of the curvature tensor field is obtained as follows. Let $X, Y, Z, W \in \mathfrak{m}$. By (4.5) we have

$$(\Delta R)(X^*,Y^*;Z^*)\cdot W^* = \Delta_{Z^*}\cdot(R(X^*,Y^*)\cdot W^*) - R(X^*,Y^*)(\Delta_{Z^*}\cdot W^*) - R(\Delta_{Z^*}\cdot X^*,Y^*)\cdot W^* - R(X^*,\Delta_{Z^*}\cdot Y^*)\cdot W^*.$$

We shall find the value at p_0. Since R is an invariant tensor field, we have $(R(X^*,Y^*)\cdot W^*)_{\tau(c)p_0} = (\tau(c))'\cdot(R(X^*,Y^*)\cdot W^*)_{p_0}$. If we denote by Δ_Z the mapping $X\in\mathfrak{m}\to\Delta_Z\cdot X = \alpha(Z,X)$ of $\mathfrak{m}$ into itself, we find $(\Delta_{Z^*}\cdot R(X^*,Y^*)\cdot W^*)_{p_0} = \Delta_Z\cdot(R(X,Y)\cdot W)$. Using the property of R that $R(fX^*,gY^*)\cdot hZ^* = fgh\,R(X^*,Y^*)\cdot Z^*$ for any functions f, g, and h, we easily find $(R(X^*,Y^*)\cdot(\Delta_{Z^*}\cdot W^*))_{p_0} = R(X,Y)\cdot(\Delta_Z\cdot W)$ and similar expressions for the other terms. Therefore we obtain the following expression of the value at p_0 of the covariant differential ΔR:

$$(9.7)\quad (\Delta R)(X,Y;Z)\cdot W = \Delta_Z\cdot(R(X,Y)\cdot W) - R(X,Y)(\Delta_Z\cdot W) - R(\Delta_Z\cdot X,Y)\cdot W - R(X,\Delta_Z\cdot Y)\cdot W.$$

10. Canonical affine connections. We shall consider the following properties of invariant affine connection on G/H.

(A1) Let $x(s)$ be the 1-parameter subgroup of G generated by an element $X\in\mathfrak{m}$, and let $x^*(s)$ be the image of $x(s)$ by the projection π of G onto G/H. Then $x^*(s)$ is a path.

(A2) Let $x(s)$ and $x^*(s)$ be defined as above. Let Y be an element of $\mathfrak{m}$. Then the parallel displacement of the tangent vector Y at p_0 along the curve $x^*(s)$ is the same as the translation of Y by the 1-parameter subgroup $x(s)$.

If (A1) is satisfied, every path through p_0 is the image by π of the 1-parameter subgroup generated by a certain element of $\mathfrak{m}$. The affine connection being invariant, any path in G/H is either such a path or its translation by an element of G. Therefore, if an invariant affine connection satisfies (A1), then the affine connection is complete in the sense of Section 5.

Since the tangent vectors to the curve $x^*(s)$ are equal to $(X^*)_{x^*(s)} = (\tau(x(s))'\cdot X$, we see that (A2) implies (A1). We shall now find the condition for a connection function α on $\mathfrak{m}\times\mathfrak{m}$ to define an invariant affine connection which satisfies (A1) and (A2) respectively. First we prove

(10.1) *Let $x(s)$ and $x^*(s)$ be as in the assumption of* (A1). *Then $(t(Y^*)(X^*)) = (\alpha(X, Y))^*$ on $x^*(s)$ for every value of s such that $x(s) \in N$.*

We may assume that X and Y belong to a base $X_1, X_2, \cdots, X_m$ of $\mathfrak{m}$, say $X = X_1$ and $Y = X_2$. Let s_1 be any fixed value such that $x(s_1) \in N$. We set $(\tau(x(-s_1)))' \cdot X_i^* = \sum_{\mu=1}^{m} \phi_{\mu i} X_\mu^*$ in a small neighborhood of p_0. Since $x(s)$ is a 1-parameter subgroup, we have for sufficiently small s

$$\begin{aligned}((\tau(x(-s_1)))' \cdot X_i^*)_{x^*(s)} &= (\tau(x(-s_1))' \cdot (X_i^*)_{x^*(s+s_1)} \\ &= (\tau(x(-s_1))) \cdot (\tau(x(s+s_1)))' \cdot X_i = (\tau(x(s)))' \cdot X_i = (X_i^*)_{x^*(s)},\end{aligned}$$

which implies $(X_i^*)_{x^*(s)} = \sum_{\mu=1}^{m} \phi_{\mu i}(x^*(s))(X_\mu^*)_{x^*(s)}$, that is $\phi_{\mu i}(x^*(s)) = \delta_{\mu i}$ for sufficiently small s. In particular, we have $\phi_{\mu i}(p_0) = \delta_{\mu i}$. Since $(X_1^* f)_{p_0} = \lim_{s \to 0} (1/s)(f(x^*(s)) - f(p_0))$ for any function f in N^*, we have $(X_1^* \phi_{\mu i})_{p_0} = 0$. Hence we get

$$\begin{aligned}(t(X_2^*)(X_1^*))_{x^*(s)} &= (\tau(x(s)))' \cdot (t^*(\sum_{\mu=1}^{m} \phi_{\mu 2} X_\mu^*)(\sum_{\nu=1}^{m} \phi_{\nu 1} X_\nu^*))_{p_0} \\ &= (\tau(x(s)))' \cdot (\sum_{\mu,\nu=1}^{m} \delta_{\mu 2}\delta_{\nu 1}(t^*(X_\mu^*)(X_\nu^*))_{p_0}) \\ &\quad + (\tau(x(s)))' \cdot (\sum_{\mu,\nu=1}^{m} \phi_{\nu 1}(p_0)(X_\nu^* \phi_{\mu 2})_{p_0} X_\mu) \\ &= (\tau(x(s)))' \cdot \alpha(X_1, X_2) = (\alpha(X_1, X_2))^*_{x^*(s)},\end{aligned}$$

which is the desired result.

From (10.1) and from our formulation of the parallel displacement (5.1) we see that

(10.2) *The invariant affine connection defined by α satisfies* (A1) *if and only if $\alpha(X, X) = 0$ for every $X \in \mathfrak{m}$.*

(10.3) *The invariant affine connection defined by α satisfies* (A2) *if and only if $\alpha(X, Y) = 0$ on $\mathfrak{m} \times \mathfrak{m}$.*

From (9.1) and (10.2) we have

THEOREM 10.1. *On a reductive homogeneous space G/H, there exists one and only one invariant affine connection which has a trivial torsion tensor and which satisfies* (A1). *It is defined by the connection function $\alpha(X, Y) = (1/2)[X, Y]_{\mathfrak{m}}$ on $\mathfrak{m} \times \mathfrak{m}$.*

We shall call this invariant affine connection the *canonical affine con-*

nection of the first kind on G/H with respect to a fixed decomposition of the Lie algebra $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$.

We have from (10.3)

THEOREM 10.2. *On a reductive homogeneous space G/H, there exists one and only one invariant affine connection which satisfies* (A2). *It is defined by the connection function which is identically zero on $\mathfrak{m} \times \mathfrak{m}$.*

We shall call this invariant affine connection the *canonical affine connection of the second kind* on G/H with respect to a fixed decomposition of the Lie algebra $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$.

THEOREM 10.3. *The canonical affine connection of the second kind on a reductive homogeneous space G/H has the following torsion and curvature tensors:*

$$T(X, Y) = -[X, Y]_{\mathfrak{m}}, \qquad R(X, Y) \cdot Z = -[[X, Y]_{\mathfrak{h}}, Z].$$

The covariant differentials of T and R are both zero.

This follows from those expression for T, R and ΔR which have been given in Section 9. $\Delta T = 0$ is also easy to prove.

11. Affine connections on a group space. We shall clarify E. Cartan's study of affine connections on the space of a connected Lie group.[5] Let G be a connected Lie group with Lie algebra $\mathfrak{g}$. The three affine connections which were studied by Cartan are the following.

a) (—)-connection: A left invariant affine connection on G which satisfies the condition (A2); it is defined by the connection function $\alpha = 0$ on $\mathfrak{g} \times \mathfrak{g}$. $T(X, Y) = -[X, Y]$ and $R = 0$.

b) (+)-connection: A right invariant affine connection which satisfies the condition (A2) in the sense of right translations, that is, the parallel displacement along a 1-parameter subgroup $x(s)$ is the same as the right translation by $x(s)$. If we consider $\mathfrak{g}$ as the Lie algebra of left invariant infinitesimal transformations on G (as we do for the case a)), then this connection is defined by the connection function $\alpha(X, Y) = [X, Y]$ and $T(X, Y) = [X, Y]$, $R = 0$.

c) (0)-connection: A left invariant affine connection which satisfies the condition (A1) and which has a trivial torsion tensor; it is given by $\alpha(X, Y) = (1/2)[X, Y]$. $R(X, Y) = -(1/4)ad([X, Y])$.

[5] See [2] and Chapter V of [8].

All these connections have 1-parameter subgroups as paths. They are all bi-invariant, that is, invariant by right as well as left translations of G, since their connection functions are invariant by $ad(G)$ in $\mathfrak{g}$.

12. Holonomy groups. Let G/H be a reductive homogeneous space with a fixed decomposition of the Lie algebra $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$. We first prove

(12.1) *Let $\mathfrak{h}_1$ be the subspace of $\mathfrak{h}$ spanned by all elements of the form $[X, Y]_{\mathfrak{h}}$, $X, Y \in \mathfrak{m}$. Then $\mathfrak{h}_1$ is an ideal of $\mathfrak{h}$.*

In fact, for any $X, Y \in \mathfrak{m}$ and $U \in \mathfrak{h}$ we have

$$[U, [X, Y]] = [U, [X, Y]_{\mathfrak{m}}] + [U, [X, Y]_{\mathfrak{h}}],$$

where $[U, [X, Y]_{\mathfrak{m}}] \in \mathfrak{m}$ and $[U, [X, Y]_{\mathfrak{h}}] \in \mathfrak{h}$. On the other hand, the Jacobian identity shows that

$$\begin{aligned}[U, [X, Y]] &= [X, [U, Y]] + [[U, X], Y] \\ &= [X, [U, Y]]_{\mathfrak{h}} + [X, [U, Y]]_{\mathfrak{m}} + [[U, X], Y]_{\mathfrak{h}} + [[U, X], Y]_{\mathfrak{m}}.\end{aligned}$$

Hence $[U, [X, Y]_{\mathfrak{h}}] = [X, [U, Y]]_{\mathfrak{h}} + [[U, X], Y]_{\mathfrak{h}} \in \mathfrak{h}_1$, which proves that $\mathfrak{h}_1$ is an ideal of $\mathfrak{h}$.

We have from a special case of the theorem of Chevalley stated in Section 6 and from Theorem 10.3.

THEOREM 12.1. *Let G/H be a reductive homogeneous space. The Lie algebra of the restricted holonomy group of the canonical affine connection of the second kind is isomorphic with $ad(\mathfrak{h}_1)$ acting on $\mathfrak{m}$, where $\mathfrak{h}_1$ is the ideal of $\mathfrak{h}$ spanned by all elements $[X, Y]_{\mathfrak{h}}$, $X, Y \in \mathfrak{m}$. The restricted holonomy group is isomorphic with $ad(H_1)$ acting on $\mathfrak{m}$, where H_1 is the invariant Lie subgroup of H_0 generated by the ideal $\mathfrak{h}_1$.*

This is a generalization of a well known theorem concerning riemannian symmetric spaces.[6]

It is easy to prove

THEOREM 12.2. *Let D be a discrete subgroup of a connected Lie group G. The holonomy group of the canonical affine connection of the second kind on G/D is isomorphic with $ad(D)$ acting on the Lie algebra $\mathfrak{g}$.*

13. Metric connection. It is well known that, given a symmetric non-degenerate tensor field of type (0, 2), that is, the so-called (positive-definite

[6] See [1], [5] and [10].

or indefinite) metric ds^2, there exists a unique affine connection without torsion whose parallelism leaves the given ds^2 invariant.

Let G/H be a reductive homogeneous space with a fixed decomposition of the Lie algebra $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$. Let B be a real-valued non-degenerate symmetric bilinear function on $\mathfrak{m} \times \mathfrak{m}$ which is invariant by $ad(H)$, that is, $B(ad(h)X, ad(h)Y) = B(X, Y)$ for $X, Y \in \mathfrak{m}$ and $h \in H$. B determines an invariant (indefinite) metric B^* on G/H. We shall find the connection function α which defines the invariant affine connection arising from the metric.

Let X, Y and Z be in $\mathfrak{m}$. Let $z(s)$ be the 1-parameter subgroup generated by Z and $z^*(s)$ the image of $z(s)$ by π. Denote by $X^\dagger$ and $Y^\dagger$ the results of the parallel displacement of the tangent vectors X and Y, respectively, along the curve $z^*(s)$. Taking a base $X_1, X_2, \cdots, X_m$ of $\mathfrak{m}$, we write $X^\dagger = \sum_{i=1}^{m} \phi_i X_i{}^*$ and $Y^\dagger = \sum_{i=1}^{m} \psi_i X_i{}^*$ where ϕ_i and ψ_i are differentiable functions such that $X = \sum_{i=1}^{m} \phi_i(p_0) X_i$ and $Y = \sum_{i=1}^{m} \psi_i(p_0) X_i$. Since the parallel displacement leaves B^* invariant, we have

$$B^*(X^\dagger, Y^\dagger) = \sum_{i,j=1}^{m} \phi_i \psi_i B(X_i{}^*, X_j{}^*) = B(X, Y) = \sum_{i,j=1}^{m} \phi_i(p_0)\psi_j(p_0) B(X_i, X_j).$$

B^* being invariant by the transformations of G, we have $B^*(X_i{}^*, X_j{}^*) = B(X_i, X_j)$ and hence $\sum_{i,j=1}^{m} \phi_i \psi_j B(X_i, X_j) = \sum_{i,j=1}^{m} \phi_i(p_0)\psi_j(p_0) B(X_i, X_j)$. We apply $(Z^*)_{p_0}$ to both sides and get

$$\sum_{i,j=1}^{m} \{ (Z^*\phi_i)_{p_0}\psi_j(p_0) + \phi_i(p_0)\,(Z^*\psi_j)_{p_0} \} B(X_i, X_j) = 0.$$

Since $(Z^*\phi_i)_{p_0} X_i = -\alpha(Z, X)$ as is easily seen, we get

$$B(\alpha(Z, X), Y) + B(X, \alpha(Z, Y)) = 0.$$

We introduce $U(X, Y) = \alpha(X, Y) - (1/2)[X, Y]_{\mathfrak{m}}$. Since the affine connection has a trivial torsion tensor, we have $\alpha(X, Y) - \alpha(Y, X) = [X, Y]_{\mathfrak{m}}$ and hence $U(X, Y) = U(Y, X)$. Thus we get

$$(1/2)\{B([X, Z]_{\mathfrak{m}}, Y) + B([Y, Z]_{\mathfrak{m}}, X)\} = B(U(Z, X), Y) + B(X, U(Z, Y)).$$

From this and the other two equalities obtained by the cyclic permutation of X, Y and Z, we finally obtain

$$B(U(X, Y), Z) = (1/2)\{B([X, Z]_{\mathfrak{m}}, Y) + B(X, [Y, Z]_{\mathfrak{m}})\}. \tag{13.1}$$

This equality determines $U(X, Y)$ uniquely for given X and Y since B is non-degenerate.

THEOREM 13.1. *Let G/H be a reductive homogeneous space with a fixed decomposition of the Lie algebra $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$. Let B^* be an invariant (indefinite) metric on G/H and B the bilinear function on $\mathfrak{m} \times \mathfrak{m}$ corresponding to B^*. Then the connection function of the affine connection induced by the metric B^* is*

$$\alpha(X, Y) = (1/2)[X, Y]_{\mathfrak{m}} + U(X, Y),$$

where $U(X, Y)$ is determined by (13.1).

In order for B to induce the canonical affine connection of the first kind, it is necessary and sufficient that $U(X, Y)$ is identically zero, that is,

$$B([X, Z]_{\mathfrak{m}}, Y) + B(X, [Y, Z]_{\mathfrak{m}}) = 0, \qquad X, Y, Z \in \mathfrak{m}. \tag{13.2}$$

In particular, we have

THEOREM 13.2. *Let G be a compact connected Lie group and H a closed subgroup. Let B be a positive definite quadratic form on the Lie algebra $\mathfrak{g}$ which is invariant by $ad(G)$. Let $\mathfrak{m}$ be the orthogonal subspace to the subalgebra $\mathfrak{h}$ of H. Then the riemannian metric on G/H determined by the restriction of B to $\mathfrak{m} \times \mathfrak{m}$ induces the canonical affine connection of the first kind on G/H.*

Proof. B is invariant by $ad(G)$ and hence its restriction to $\mathfrak{m} \times \mathfrak{m}$ satisfies (13.2).

Chapter III. Symmetric Affine Spaces.

14. Symmetric homogeneous spaces.[7] Let G be a connected Lie group and σ an involutive automorphism of G, that is, $\sigma^2 =$ the identity automorphism. The totality of fixed points H_σ of σ is a closed subgroup of G. A homogeneous space G/H is called a *symmetric homogeneous space* (with respect to σ) if H lies between H_σ and the identity component of H_σ. In certain cases it will be necessary to assume that G is effective on G/H as a transformation group.

Let G/H be a symmetric homogeneous space defined by σ. The *sym-*

[7] See [3]. In this Chapter, I have borrowed some of the ideas presented by Dr. Koszul at the seminar mentioned in the Introduction.

metry at the point p_0 is an involutive differentiable homeomorphism of G/H onto itself which is defined by

$$\sigma_0 \cdot (p) = \tau(a^\sigma) \cdot p_0 \qquad \text{for } p = \tau(a)p_0,\ a \in G, \tag{14.1}$$

the right hand side being independent of the choice of $a \in G$ such that $p = \tau(a) \cdot p_0$. A *transvection* is an element of G such that $a^\sigma = a^{-1}$. We define a mapping ρ of G/H into the set of transvections by $\rho(p) = aa^{-\sigma}$, the right hand side again being independent of the choice of $a \in G$ such that $p = \tau(a) \cdot p_0$. If p is the image of a transvection $a \in G$ by π, then $\rho(p) = a^2$. The symmetry at any point p is defined as the mapping $\tau(\rho(p)) \cdot \sigma_0$ of G/H onto itself.

A symmetric homogeneous space G/H is always reductive, and we take the following canonical decomposition of the Lie algebra. Let $\mathfrak{m}$ be the eigen-space for the eigen-value -1 of the involutive automorphism, denoted by the same σ, of the Lie algebra $\mathfrak{g}$ which is induced by the automorphism σ of G. Then $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$, where $\mathfrak{h}$ is the subalgebra of the identity component H_0 of H. We have $X^\sigma = -X$ for $X \in \mathfrak{m}$, $U^\sigma = U$ for $U \in \mathfrak{h}$ and $ad(H)\mathfrak{m} \subset \mathfrak{m}$. The 1-parameter subgroup generated by any $X \in \mathfrak{m}$ consists of transvections: $(x(s))^\sigma = (x(s))^{-1} = x(-s)$. If we take a neighborhood V of e with canonical coordinates $x^1, x^2, \cdots, x^m, x^{m+1}, \cdots, x^n$ with respect to $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$ as we did in Section 7, then any point with coordinates such that $x^{m+1} = x^{m+2} = \cdots x^n = 0$ is a transvection. For a point p with coordinates $(x^1, x^2, \cdots, x^m)$ near p_0 in G/H, the point $\sigma_0 \cdot p$ has coordinates $(-x^1, -x^2, \cdots, -x^m)$. We see that p_0 is an isolated fixed point of σ_0.

From now on we shall always use the canonical decomposition $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$ for a symmetric homogeneous space. We have

$$[X, Y] \in \mathfrak{h} \text{ for any } X, Y \in \mathfrak{m}. \tag{14.2}$$

Keeping the notation in Section 7 and using the fact that σ_0 commutes with $\tau(a)$ for every $a \in G$, we can easily prove

$$(\sigma_0)' \cdot X^* = -X^* \text{ for any } X \in \mathfrak{m}. \tag{14.3}$$

15. Canonical affine connection on a symmetric homogeneous space. By the results of Chapter II we can clarify the existence and properties of a natural affine connection on any symmetric space. From Theorem 10.1, Theorem 10.2 and (14.2) we have immediately

THEOREM 15.1. *The canonical affine connection of the second kind of a symmetric homogeneous space G/H has a trivial torsion tensor. In other words, the canonical affine connections of the first and second kind coincide.*

We shall call this invariant affine connection the *canonical affine connection* of a symmetric homogeneous space.

Conversely, we have

THEOREM 15.2. *Let G/H be a reductive homogeneous space. If the canonical affine connections of the first and second kind (with respect to a fixed decomposition $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$) coincide, then $[\mathfrak{m}, \mathfrak{m}] \subset \mathfrak{h}$. If moreover G/H is simply connected, it can be considered as a symmetric homogeneous space $G^\dagger/H^\dagger$ where $G^\dagger$ is the simply connected covering group of G and $H^\dagger$ is the subgroup of $G^\dagger$ generated by $\mathfrak{h}$.*

Proof. The first assertion is already clear. The linear endomorphism σ of $\mathfrak{g}$ defined by $X^\sigma = -X$ for $X \in \mathfrak{m}$ and $U^\sigma = U$ for $U \in \mathfrak{h}$ is an involutive automorphism of the Lie algebra g. Let $\sigma^\dagger$ be the involutive automorphism of $G^\dagger$ induced by σ. It is clear that $H^\dagger$ is contained in the set of fixed points of σ so that $G^\dagger/H^\dagger$ is a symmetric space defined by $\sigma^\dagger$. Since G/H is simply connected, $G/H = G^\dagger/H^\dagger$.

THEOREM 15.3. *Let G/H be a symmetric homogeneous space. The canonical affine connection is the only invariant affine connection on G/H which is invariant by the symmetry at each point.*

Proof. Let α be a connection function on $\mathfrak{m} \times \mathfrak{m}$. We shall prove that the invariant affine connection t defined by α is invariant by the symmetry at each point if and only if α is identically zero.

Suppose first that t is invariant by σ_0. Then we have

$$(\sigma_0)' \cdot (t(Y^*)(X^*) = t((\sigma_0)' \cdot Y^*)((\sigma_0)' \cdot X^*)$$

for any $X, Y \in \mathfrak{m}$. Using (14.3) we see that $\alpha(X, Y) = 0$.

Conversely, we prove that the canonical affine connection is invariant by the symmetry at each point. It is sufficient to prove the invariance by σ_0. Let t' be the affine connection defined by

$$t'(X^\dagger)(Y^\dagger) = (\sigma_0)' \cdot (t((\sigma_0)' \cdot X^\dagger)((\sigma_0)' \cdot Y^\dagger))$$

for any infinitesimal transformations $X^\dagger$ and $Y^\dagger$ on G/H. Since σ_0 commutes with every $\tau(a)$, $a \in G$, we see that t' is an invariant affine connection. We have $(t'(X^*)(Y^*))_{p_0} = 0$ in virtue of (14.3). Since an invariant affine connection is uniquely determined by the corresponding connection function, we see that t and t' coincide, that is, t is invariant by σ_0.

THEOREM 15.4. *Let G/H be a symmetric homogeneous space with its*

canonical affine connection. The covariant differential of any invariant tensor field on G/H is zero. In particular, every invariant differential form is closed.

Proof. Let F be an invariant tensor field of type (i, j). Since the covariant differentiation commutes with $\tau(a)$, $a \in G$, ΔF is also an invariant tensor field of type $(i, j+1)$. Δ also commutes with σ_0. Therefore we have at p_0

$$(-1)^{i+j+1}\Delta F = \sigma_0 \cdot \Delta \cdot F = \Delta \cdot \sigma_0 \cdot F = (-1)^{i+j}\Delta F$$

and hence $\Delta F = 0$. The second assertion follows from the fact that the exterior derivative $d\omega$ of a differential form ω is the alternation of the covariant differential $\Delta\omega$ of ω in the case where the affine connection has a trivial torsion tensor.

THEOREM 15. 5. *If G/H is a symmetric homogeneous space with semi-simple G, the canonical affine connection is induced by an invariant (indefinite) metric on G/H.*

Proof. Let $\phi(X, Y) =$ Trace of $ad(X)ad(Y)$ be the fundamental bilinear form of $\mathfrak{g}$. Since it is invariant by every automorphism of $\mathfrak{g}$, we have in particular $\phi(X, U) = \phi(X^\sigma, U^\sigma) = \phi(-X, U)$ and hence $\phi(X, U) = 0$ for any $X \in \mathfrak{m}$ and $U \in \mathfrak{h}$. Let $B(X, Y)$ be the restriction of $\phi(X, Y)$ to $\mathfrak{m} \times \mathfrak{m}$. It is clearly invariant by $ad(H)$. It is non-degenerate; in fact, if $B(X, Y) = 0$ for all $Y \in \mathfrak{m}$, then $\phi(X, Y)$ for all $Y \in \mathfrak{g}$ and hence $X = 0$ since ϕ is non-degenerate. $B(X, Y)$ satisfies (13. 2). The remark which follows Theorem 13. 1 shows that the canonical affine connection is a metric connection.

THEOREM 15. 6. *Let G/H be a symmetric homogeneous space. An invariant (indefinite) metric on G/H, if there exists any, induces the canonical affine connection.*

Proof. Let B^* be an invariant metric on G/H. Since it is an invariant tensor field of type $(0, 2)$, we easily see that it is invariant by σ_0. Therefore the induced affine connection is invariant by σ_0 and coincides with the canonical affine connection in virtue of Theorem 15. 3.

In particular, if H is compact, G/H admits an invariant positive-definite metric and any such metric induces the canonical affine connection. This is what is called a *riemannian symmetric homogeneous space.*

16. Irreducible symmetric homogeneous spaces. Let G/H be a symmetric homogeneous space with the canonical decomposition of the Lie algebra

$\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$. If $ad(\mathfrak{h})$ in $\mathfrak{m}$ is irreducible, G/H is called *irreducible*. In this section, we assume that G is effective on G/H as a transformation group. It follows then that $\mathfrak{h}$ does not contain any non-trivial ideal of $\mathfrak{g}$. This last condition is equivalent to the condition that the representation $\mathfrak{h} \to ad(\mathfrak{h})$ in $\mathfrak{m}$ is faithful.

We shall first prove

(16. 1) *If G/H is irreducible, then either $\mathfrak{g}$ is semi-simple or $[\mathfrak{m}, \mathfrak{m}] = 0$.*

In order to prove this, let $\phi(X, Y)$ be the fundamental bilinear form of $\mathfrak{g}$. We have $\phi(X, U) = 0$ for any $X \in \mathfrak{m}$ and $U \in \mathfrak{h}$ as we have seen before. Let $\mathfrak{r} = \{X \in \mathfrak{g} \mid \phi(X, Y) = 0 \text{ for all } Y \in \mathfrak{g}\}$. $\mathfrak{r}$ is an ideal of $\mathfrak{g}$ which is contained in the radical of $\mathfrak{g}$. $\mathfrak{r} \cap \mathfrak{m}$ is a subspace of $\mathfrak{m}$ which is invariant by $ad(\mathfrak{h})$. $\mathfrak{m}$ being irreducible by assumption, we have either $\mathfrak{r} \cap \mathfrak{m} = (0)$ or $\mathfrak{r} \cap \mathfrak{m} = \mathfrak{m}$.

Let $\mathfrak{r} \cap \mathfrak{m} = (0)$. We shall prove that ϕ is non-degenerate, that is, $\mathfrak{g}$ is semi-simple. Let $X = Y + U$, where $Y \in \mathfrak{m}$ and $U \in \mathfrak{h}$, be in $\mathfrak{r}$. For any $Z \in \mathfrak{m}$, we have $\phi(X, Z) = \phi(Y, Z) + \phi(U, Z)$ and hence $\phi(Y, Z) = 0$. Therefore $Y \in \mathfrak{r} \cap \mathfrak{m}$ and $Y = 0$. We have $X = U \in \mathfrak{h} \cap \mathfrak{r}$. For any $Z \in \mathfrak{m}$, we have $[U, Z] \in \mathfrak{m} \cap \mathfrak{r} = (0)$ which means that $[U, \mathfrak{m}] = 0$. $ad(\mathfrak{h})$ in $\mathfrak{m}$ being a faithful representation of $\mathfrak{h}$, we must have $U = 0$. We have thereby proved that $X = 0$ and that ϕ is non-degenerate.

Next we consider the case where $\mathfrak{r} \cap \mathfrak{m} = \mathfrak{m}$, that is, $\mathfrak{m} \subset \mathfrak{r}$. We remark that $[\mathfrak{h}, \mathfrak{r}]$ is contained in the nilpotent radical $\mathfrak{n}$ of $\mathfrak{g}$ since any derivation of $\mathfrak{g}$ maps the radical of $\mathfrak{g}$, and a fortiori $\mathfrak{r}$, into $\mathfrak{n}$. $\mathfrak{m}$ being irreducible, we have $[\mathfrak{h}, \mathfrak{m}] = \mathfrak{m}$ and hence $\mathfrak{m} \subset [\mathfrak{h}, \mathfrak{r}] \subset \mathfrak{n}$. Since $\mathfrak{n}$ is nilpotent, there exists a positive integer k such that $\mathfrak{n}^{(k)} = [\mathfrak{n}, \mathfrak{n}^{(k-1)}] = (0)$ and hence $\mathfrak{m}^{(k)} = (0)$. If k is even, then $\mathfrak{m}^{(k-1)}$ is a subspace of $\mathfrak{m}$ which is invariant by $ad(\mathfrak{h})$ which means that either $\mathfrak{m}^{(k-1)} = \mathfrak{m}$ or $\mathfrak{m}^{(k-1)} = (0)$. In the first case we have $[\mathfrak{m}, \mathfrak{m}] = [\mathfrak{m}, \mathfrak{m}^{(k-1)}] = \mathfrak{m}^{(k)} = (0)$. In the second case, we have $\mathfrak{m}^{(k-1)} = [\mathfrak{m}, \mathfrak{m}^{(k-2)}]$ where $\mathfrak{m}^{(k-2)}$ is a subspace of $\mathfrak{h}$ since $k - 2$ is even. $ad(h)$ being faithful, we must have $\mathfrak{m}^{(k-2)} = (0)$. This argument shows that $[\mathfrak{m}, \mathfrak{m}] = (0)$.

Next we prove

(16. 2) *If $\mathfrak{g}$ is semi-simple, then $[\mathfrak{m}, \mathfrak{m}] = \mathfrak{h}$.*

Let $\mathfrak{h}_1 = [\mathfrak{m}, \mathfrak{m}]$. It is straightforward to see that $\mathfrak{h}_1 \neq (0)$ is an ideal of $\mathfrak{h}$ and that $\mathfrak{m} + \mathfrak{h}_1$ is an ideal of $\mathfrak{g}$. $\mathfrak{g}$ being semi-simple, there exists a supplementary ideal $\mathfrak{k}$ such that $\mathfrak{g}$ is the direct sum of $\mathfrak{m} + \mathfrak{h}_1$ and $\mathfrak{k}$. It is sufficient

to prove that $\mathfrak{k}$ is contained in $\mathfrak{h}$; then $\mathfrak{k}$ must be zero since $\mathfrak{h}$ does not contain any non-trivial ideal of $\mathfrak{g}$. Let $X = Y + U$ where $Y \in \mathfrak{m}$, $U \in \mathfrak{h}$, be any element of $\mathfrak{k}$. We shall prove that $Y = 0$. For any $Z \in \mathfrak{m}$, we have $[Z, X] = [Z, Y] + [Z, U] = 0$, where $[Z, Y] \in \mathfrak{h}_1$ and $[Z, U] \in \mathfrak{m}$. Hence $[Z, Y] = [Z, U] = 0$. This shows that $[Y, \mathfrak{m}] = 0$. Next, we have for any $V \in \mathfrak{h}_1$, $[V, X] = [V, Y] + [V, U] = 0$ and hence $[V, Y] = 0$, which shows that $[Y, \mathfrak{h}_1] = 0$. Finally $[Y, \mathfrak{k}] = 0$ is obvious since $Y \in \mathfrak{m}$. Thus $[Y, \mathfrak{g}] = 0$ and $Y = 0$ since $\mathfrak{g}$ is semi-simple. We have completed the proof of (16.2).

If $[\mathfrak{m}, \mathfrak{m}] = (0)$, the canonical affine connection of G/H is locally flat.

We now prove the following theorem which is an extension of a theorem of Cartan in the riemannian case.

THEOREM 16.1. *Let G/H be an irreducible symmetric homogeneous space where G is semi-simple and effective on G/H. Then the largest connected group of affine transformations of G/H (of course, with respect to the canonical affine connection) is equal to G.*

Proof. Since G/H is complete, the group of all affine transformations of G/H is a Lie group.[8] Let $G^\dagger$ be the identity component of this group and let $G^\dagger/H^\dagger = G/H$. We shall show that $G^\dagger/H^\dagger$ can be considered as a symmetric homogeneous space. Let σ_0 be the symmetry as p_0 and define an involutive automorphism $\sigma^\dagger$ of $G^\dagger$ as follows: $(a^\dagger)^{\sigma^\dagger} = \sigma_0 \cdot a^\dagger \cdot \sigma_0$, which is again an affine transformation since the canonical affine connection is invariant by σ_0. $\sigma^\dagger$ so defined is obviously an extension of the automorphism σ which defines G/H. Let $K^\dagger$ be the subgroup of fixed points of $\sigma^\dagger$. We show that $H^\dagger$ lies between $K^\dagger$ and its identity component $K^\dagger_0$.

Let $a^\dagger$ be any element of $H^\dagger$. Since $a^\dagger \cdot p_0 = p_0$, $a^\dagger$ induces a linear transformation of the tangent space $\mathfrak{m}$ at p_0, and so does $(a^\dagger)^{\sigma^\dagger}$. These two linear transformations of $\mathfrak{m}$ coincide since $(\sigma_0)' \cdot X = -X$ for $X \in \mathfrak{m}$. $a^\dagger$ and $(a^\dagger)^{\sigma^\dagger}$ being affine transformations, they coincide.[9] This proves that $H^\dagger \subset K^\dagger$. In order to prove that $K^\dagger_0 \subset H^\dagger$, it is sufficient to prove that every element in a sufficiently small neighborhood of e in $K^\dagger$ lies in $H^\dagger$. Let N^* be a neighborhood of p_0 in which p_0 is the only fixed point of σ_0, and let U be a neighborhood of e in $G^\dagger$ such that $a^\dagger \cdot p_0 \in N$ for every $a \in U$. Let $a^\dagger \in U \cap K^\dagger$. Since $(a^\dagger)\sigma^\dagger = a^\dagger$, we have $a^\dagger \cdot p_0 = \sigma_0 \cdot a^\dagger \cdot \sigma_0 \cdot p = \sigma_0 \cdot a^\dagger \cdot p_0$, which implies

[8] See [9].
[9] See [9].

that $a^\dagger \cdot p_0 = p_0$ since $a^\dagger \cdot p_0 \in N^*$ and since p_0 is the only fixed point of σ_0 in N^*. This proves that $K^\dagger_0 \subset H$.

Let $\mathfrak{g}^\dagger = \mathfrak{m}^\dagger + \mathfrak{h}^\dagger$ be the canonical decomposition of the Lie algebra of the symmetric homogeneous space $G^\dagger/H^\dagger$. Since $G/H = G^\dagger/H^\dagger$, we have $\mathfrak{m}^\dagger = \mathfrak{m}$. Since $\mathfrak{h} \subset \mathfrak{h}^\dagger$, $ad(\mathfrak{h}^\dagger)$ in $\mathfrak{m}$ is irreducible. Since $G/H = G^\dagger/H^\dagger$ is not locally flat by assumption, (16.1) and (16.2) imply that $[\mathfrak{m}, \mathfrak{m}] = \mathfrak{h}^\dagger$. On the other hand, $[\mathfrak{m}, \mathfrak{m}] = \mathfrak{h}$. Hence $\mathfrak{h}^\dagger = \mathfrak{h}$ and $\mathfrak{g}^\dagger = \mathfrak{g}$. We have thus proved that $G^\dagger = G$.

As a consequence we have

THEOREM 16.2. *Let G/H be an irreducible riemannian symmetric homogeneous space where G is semi-simple and effective on G/H. Then the largest group of affine transformations and the largest group of isometries of G/H have the same identity component.*

Proof. The identity components of these groups coincide with G.

We remark that the same conclusion has been proved for any compact riemannian space by K. Yano.[10]

17. Affine symmetric spaces. Just as in the riemannian case, we can define an *affine symmetric space* geometrically as follows. Let M be a manifold with an affine connection. For any $p \in M$, there exists a neighborhood N_p of p with normal coordinates $x^1, x^2, \cdots, x^n$. We define the *symmetry* σ_p at p by $\sigma_p(x) = -x$ for any point $x = (x^1, x^2, \cdots, x^n) \in N_p$. σ_0 is an involutive homeomorphism of N_p onto itself with p as the only fixed point. M is called an affine symmetric space if σ_p for every $p \in M$ is an affine transformation of N_p into itself. We wish to prove

THEOREM 17.1. *The following conditions about an affinely-connected manifold are mutually equivalent:*

a) *M is an affine symmetric space;*

b) *$T = 0$ and $\Delta R = 0$;*

c) *For any $p \in M$, there exists a neighborhood N of p and an affine transformation of N onto a neighborhood of a certain symmetric homogeneous space with its canonical affine connection (we say simply that M is locally representable as a symmetric homogeneous space).*

---

[10] He proved that any 1-parameter subgroup of affine transformations of a compact orientable Riemannian manifold is in fact a 1-parameter subgroup of isometries. See [12].

The equivalence of these conditions in the riemannian case has been shown by Cartan. The affine case has been partly treated by him and J. H. C. Whitehead.[11]

Proof. Theorem 15.3 shows that c) implies a). We shall prove here that a) implies b). The proof of the most essential part b) $\rightarrow$ c) will be given as a special case of Theorem 18.1 in Chapter IV.

In order to prove that a) implies b), let p be any point of M with a normal neighborhood N_p. For each tangent vector X at p, we define an infinitesimal transformation X^* in N_p by the parallel displacement along paths, that is, if $q \in N_p$ lies on a path C joining p and q in N_p, we define $(X^*)_q$ as the result of the parallel displacement of X along C. Since $(\sigma_p)' \cdot X = -X$ at p and since σ_p is an affine transformation of N_p into itself by assumption, we easily see that $(\sigma_p)' \cdot X^* = -X^*$. The torsion tensor field T is invariant by σ_p, that is,

$$(\sigma_p)' \cdot T(X^*, Y^*) = T((\sigma_p)' \cdot X^*, (\sigma_p)' \cdot Y^*)$$

for the infinitesimal transformations X^* and Y^* corresponding to any tangent vectors X and Y at p. Using $(\sigma_p)' \cdot X^* = -X^*$, we obtain

$$((\sigma_p)' \cdot T(X^*, Y^*))_p = -(T(X^*, Y^*))_p$$

and, on the other hand,

$$(T((\sigma_p)' \cdot X^*, (\sigma_p)' \cdot Y^*))_p = (T(-X^*, -Y^*))_p = (T(X^*, Y^*))_p,$$

that is, $(T(X^*, Y^*))_p = 0$. Since X and Y are arbitrary tangent vectors at any arbitrary point p of M, we have thus proved that T is zero everywhere. $\Delta R = 0$ can be proved in a similar fashion by making use of (4.5).

Chapter IV. Affine Connection Invariant By Its Parallelism.

18. Affine connection invariant by its parallelism. We shall define a type of affine connection which is the simplest next to that treated in Chapter III. Let M be an affinely-connected manifold. Let C be a path from a point p_0 to a nearby point p_1. We denote by $(\tau(C))'$ the parallel displacement of vectors along C, which is a linear mapping of the tangent space $T(p_0)$ onto the tangent space $T(p_1)$. Let N_1 be a normal neighborhood of p_1 and let N_0 be a sufficiently small normal neighborhood of p_0. Each

[11] See [2] and [11].

point $q_0 \in N_0$ lies on a certain path with tangent vector, say X, at p_0. Let q_1 be the point which is on the path with tangent vector $(\tau(C))' \cdot X$ at p_1 and which corresponds to the same value of the canonical parameter as q_0. We denote the mapping which sends q_0 into q_1 so determined by $\tau(C)$ and call it a parallel transformation determined by the path C. $\tau(C)$ is a local homeomorhpism of N_0 into N_1 and clearly induces the linear mapping $(\tau(C))'$ of $T(p_0)$ onto $T(p_1)$.

If the affine connection of M is invariant by all parallel transformations, we say that the affine connection is *invariant by its parallelism.*

We shall now establish

THEOREM 18. 1. *The following conditions about an affinely-connected manifold are mutually equivalent*:

a) *The affine connection is invariant by its parallelism*;

b) $\Delta T = 0$ *and* $\Delta R = 0$;

c) *It is locally representable as a reductive homogeneous space G/H with its canonical affine connection of the second kind (with respect to a certain decomposition of the Lie algebra).*

Proof. Let G/H be a reductive homogeneous space with its canonical affine connection of the second kind. We see from property (A2) that a parallel transformation along a path $x^*(s)$ with origin p_0 is nothing but the transformation by the 1-parameter subgroup $x(s)$ whose projection is the given $x^*(s)$. This fact and the invariance of the affine connection by G proves that c) implies a).

a) $\rightarrow$ b) is an easy consequence of the following alternative definition of the covariant differentiation of tensor fields in any affinely-connected manifold. Let $C(s)$ be a path with tangent vector X at the origin p where s is the canonical parameter. Let $\tau(C(s))'$ denote the parallel displacement along the path from p up to the point with parameter s. If F is a tensor field of type, say $(1, 2)$, then

$$(\Delta_X \cdot F)_p(Y, Z) = \lim_{s \to 0} (1/s)\{(\tau(C(s)))' \cdot F(Y, Z) - F((\tau(C(s)))' \cdot Y, (\tau(C(s))' \cdot Z)\}$$

for any tangent vectors X, Y and Z at p.

In order to prove that b) implies c), we need some preparations, which we shall do in the next Section.

19. Certain identities derived from $\Delta T = 0$ and $\Delta R = 0$. In the following, we shall denote by X, Y and Z any infinitesimal transformations in an affinely-connected manifold M. T_X denotes the mapping $Y \to T(X, Y)$ of $\mathfrak{X}$ into itself and $\mathfrak{S}\{P(X, Y, Z)\}$ denotes the sum of the expression $P(X, Y, Z)$ over the cyclic permutations (X, Y, Z), (Y, Z, X) and (Z, X, Y). We shall derive certain identities from the conditions $\Delta T = 0$ and $\Delta R = 0$.

(19. 1) *If* $\Delta T = 0$, *then* $[\Delta_X, T_Y] = T_{\Delta_X \cdot Y}$.

Here Δ_X is the mapping $Z \to t(Z)X$ of $\mathfrak{X}$ into itself which was defined in Section 4, and $[\Delta_X, T_Y]$ is the bracket product of endomorphisms of the linear space $\mathfrak{X}$. If $\Delta T = 0$, we have

$$\begin{aligned}\Delta_Z \cdot T(X, Y) - T(\Delta_Z \cdot X, Y) - T(X, \Delta_Z \cdot Y) \\ = \Delta_Z \cdot T_X \cdot Y - T_X \cdot \Delta_Z \cdot Y - T_{\Delta_Z \cdot X} \cdot Y = 0\end{aligned}$$

which gives the desired result.

(19. 2) *If* $\Delta T = 0$, *then* $\mathfrak{S}\{R(X, Y) \cdot Z - T(T(X, Y), Z)\} = 0$.

Proof. We introduce the following affine connection t' associated to the given affine connection t:

$$t'(X)Y = t(X)Y + (1/2)T(X, Y).^{12}$$

Since the torsion tensor of the new connection t' is equal to 0, its curvature tensor R' satisfies the well known identity $\mathfrak{S}\{R'(X, Y) \cdot Z\} = 0$. Let Δ'_Y be the mapping $X \to t'(X)Y$ of $\mathfrak{X}$ into itself. From $\Delta'_Y = -(1/2)T_Y + \Delta_Y$ we obtain, by using (19. 1) and (4. 4)

$$\begin{aligned}R'(X, Y) &= [\Delta'_X, \Delta'_Y] - \Delta'_{[X,Y]} \\ &= [-(1/2)T_X + \Delta_X, -(1/2)T_Y + \Delta_Y] + (1/2)T_{[X,Y]} - \Delta_{[X,Y]} \\ &= R(X, Y) + (1/4)[T_X, T_Y] - (1/2)\{[T_X, \Delta_Y] + [\Delta_X, T_Y] - T_{[X,Y]}\} \\ &= R(X, Y) + (1/4)[T_X, T_Y] - (1/2)T_{T(X,Y)}.\end{aligned}$$

We easily see that

$$\mathfrak{S}\{[T_X, T_Y] \cdot Z\} = -2\mathfrak{S}\{T(T(X, Y), Z)\}$$

and hence we get (19. 2).

(19. 3) *If* $\Delta R = 0$, *then* $\mathfrak{S}\{R(T(X, Y), Z)\} = 0$.

Proof. Using the identity $R(\Delta_Z X, Y) - R(\Delta_Z Y, X) = [\Delta_Z, R(X, Y)]$ which is an immediate consequence of $\Delta R = 0$, we have

---

[12] If t is defined by means of Γ^k_{ij}, then t' is defined by $\Gamma'^k_{ij} = (1/2)(\Gamma^k_{ij} + \Gamma^k_{ji})$.

$$\begin{aligned}
\mathfrak{S}\{R(T(X,Y),Z)\} &= \mathfrak{S}\{R(\Delta_X \cdot Y - \Delta_Y \cdot X - [X,Y],Z) \\
&= R(\Delta_X \cdot Y, Z) - R(\Delta_Y \cdot X, Z) - R([X,Y],Z) \\
&\quad + R(\Delta_Y \cdot Z, X) - R(\Delta_Z \cdot Y, X) - R([Y,Z],X) \\
&\quad + R(\Delta_Z \cdot X, Y) - R(\Delta_X \cdot Z, Y) - R([Z,X],Y) \\
&= \mathfrak{S}\{[\Delta_Z, R(X,Y)] - R([X,Y],Z)\} \\
&= \mathfrak{S}\{[\Delta_Z, R(X,Y)] - [\Delta_{[X,Y]}, \Delta_Z] + \Delta_{[[X,Y],Z]}\} \\
&= \mathfrak{S}\{[\Delta_Z, R(X,Y) + \Delta_{[X,Y]}]\} \\
&= \mathfrak{S}\{[\Delta_Z, [\Delta_X, \Delta_Y]]\} = 0,
\end{aligned}$$

where we have used the Jacobian identity $\mathfrak{S}\{[[X,Y],Z]\} = 0$ in $\mathfrak{X}$ and the Jacobian identity $\mathfrak{S}\{[\Delta_Z, [\Delta_X, \Delta_Y]]\} = 0$ for endomorphisms.

We are now in a position to prove that b) implies c). Let M be an affinely-connected manifold such that $\Delta T = 0$ and $\Delta R = 0$. Let p be any arbitrary point of M and $\mathfrak{m}$ the tangent space at p. We denote by $\mathfrak{h}$ the Lie algebra consisting of endomorphisms of $\mathfrak{m}$ which, when extended to a derivation of the tensor algebra over $\mathfrak{m}$, leave the tensors T and R at p invariant. Since $\Delta T = 0$ and $\Delta R = 0$, Δ_X for any $X \in \mathfrak{m}$ belongs to $\mathfrak{h}$, and so does $R(X,Y) = [\Delta_X, \Delta_Y] - \Delta_{[X,Y]}$ for any X and $Y \in \mathfrak{m}$. Now we define a Lie algebra $\mathfrak{g}$ in the following fashion. As a vector space $\mathfrak{g}$ is the direct sum of $\mathfrak{m}$ and $\mathfrak{h}$: $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$, where we define the bracket operation as follows:

For $X, Y \in \mathfrak{m}$,	$[X,Y] = (-T(X,Y), -R(X,Y))$
For $X \in \mathfrak{m}$, $U \in \mathfrak{h}$,	$[X,U] = -U \cdot X$ (U operating on X)
For $U, V \in \mathfrak{h}$,	$[U,V] = UV - VU$.

The Jacobian identity is satisfied in virtue of (19.2) and (19.3), and $\mathfrak{g}$ is a Lie algebra of which $\mathfrak{h}$ is a subalgebra.

Let G be the simply connected Lie group having $\mathfrak{g}$ as the Lie algebra and H the Lie subgroup of G generated by $\mathfrak{h}$. Since $\mathfrak{h}$ consists of all endomorphisms of $\mathfrak{m}$ which leave T and R invariant, we see that H is a closed subgroup. Clearly G/H is a reductive homogeneous space with the above decomposition $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$. The canonical affine connection of the second kind on G/H has the same values of the torsion and curvature tensor fields at p_0 (the point represented by the coset H) as those of the given affine connection of M at p. Therefore we shall be able to conclude the proof of b) $\rightarrow$ c) if we prove

(19.4) *Let M and M' be two manifolds with affine connection which*

both satisfy the conditions b) *of Theorem* 18.1. *Let* $p \in M$ *and* $p' \in M'$ *and let* $\mathfrak{m}$ *and* $\mathfrak{m}'$ *be the tangent spaces at* p *and* p' *respectively. If* Λ *is a linear isomorphism of* $\mathfrak{m}$ *onto* $\mathfrak{m}'$ *which transforms the torsion and curvature tensors at* p *into those at* p', *then there exists an affine transformation of a neighborhood of* p *onto a neighborhood of* p' *which induces the given* Λ *at* p.

Or what amounts to the same,

(19.5) *An affine connection such that* $\Delta T = 0$ *and* $\Delta R = 0$ *is uniquely determined, locally, by the values of* T *and* R *at a given point.*

(19.5) is well known in the riemannian case. We shall prove the general case by a technique similar to that of Cartan.[13] Let p_0 be any point of M and $X_1, X_2, \cdots, X_n$ a frame at p_0. Let N be a normal neighborhood of p_0 and define the infinitesimal transformations $X_1^*, X_2^*, \cdots, X_n^*$ by the parallel displacement along paths as we did before. Let $t(X_j^*)(X_k^*) = \sum_{i=1}^{n} \gamma^i{}_{kj} X_i^*$ and set $\omega^i{}_p = \sum_{q=1}^{n} \gamma^i{}_{qp}\omega^q$ where $\omega^1, \omega^2, \cdots, \omega^n$ are linearly independent differential forms such that $\langle X_i, \omega^j \rangle = \delta_{ij}$. If $T(X_j^*, X_k^*) = \sum_{i=1}^{n} T^i{}_{jk} X_i^*$ is the expression for the torsion tensor field in N, then $T^i = \sum_{j,k=1}^{n} T^i{}_{jk}\omega^j \wedge \omega^k$ is, for each i, a differential form of degree 2. It is easy to verify

$$(19.6) \quad (1/2)\sum_{j,k=1}^{n} T^i{}_{jk}\omega^j \wedge \omega^k = \sum_{p=1}^{n} \omega^i{}_p \wedge \omega^p - d\omega^i \text{ for each } i.$$

If $R(X_j^*, X_k^*) \cdot X_l^* = \sum_{i=1}^{n} R^i{}_{ljk} X_i^*$ is the expression for the curvature tensor field in N, then $R^i{}_l = \sum_{j,k=1}^{n} R^i{}_{ljk}\omega^j \wedge \omega^k$ is, for each i and l, a differential form of degree 2 and

$$(19.7) \quad \sum_{j,k=1}^{n} (1/2) R^i{}_{ljk}\omega^j \wedge \omega^k = \sum_{p=1}^{n} \omega^i{}_p \wedge \omega^p{}_l - d\omega^i{}_l \text{ for each } i \text{ and } l.$$

We remark that we have used the formula

$$d\omega^i(X_j^*, X_k^*) = (1/2)\omega^i([X_j^*, X_k^*]).$$

Since $\Delta T = 0$ and $\Delta R = 0$, T and R are invariant by the parallel displacement. From the choice of X_i^*, we see that the coefficients $T^i{}_{jk}$ and $R^i{}_{ljk}$ are constants. We shall show that (19.6) and (19.7) determine ω^i

[13] See Chapter X, II, pp. 236-238 and Chapter XI, II, pp. 262-265 of [4].

and $\omega^i{}_p$ and hence the affine connection uniquely. We consider coordinates $(a^1, a^2, \cdots, a^n; t)$ instead of the normal coordinates (x^i) in N. The coordinates (a^i) remain constant along the path with tangent vector $\sum_{i=1}^{n} a^i X_i$ at p_0. It is easy to see that ω^i, when expressed in terms of a^i and t, can be written as $\omega^i(t, a^i; dt, da^i) = a^i dt + \omega^{*i}(t, a^i; da^i)$ where ω^{*i} does not contain dt. On the other hand, $\omega^i{}_p(t, a^i; dt, da^i) = \omega^{*i}{}_p(t, a^i; da^i)$ does not contain dt either. We substitute these expressions into (19. 6) and (19. 7) and take the terms which contain dt. We then obtain the following system of differential equations:

$$\partial\omega^{*i}/\partial t = da^i + \sum_{p=1}^{n} a^p \omega^{*i}{}_p + \sum_{j,k=1}^{n} T^i{}_{jk} a^j \omega^{*k} \tag{19. 8}$$

$$\partial\omega^{*i}{}_l/\partial t = \sum_{j,k=1}^{n} R^i{}_{ljk} a^k \omega^{*j},$$

with the initial conditions

$$\omega^{*i}(0, a^i; da^i) = 0, \quad \omega^{*i}{}_p(0, a^i; da^i) = 0. \tag{19. 9}$$

The differential equations (19. 8), together with the initial conditions (19. 9), determine ω^{*i} and $\omega^{*i}{}_p$ uniquely as linear expressions in $da^1, da^2, \cdots, da^n$ with coefficients which are entire functions in $t, a^1, a^2, \cdots, a^n$. Therefore ω^i and $\omega^i{}_p$ and hence the affine connection in N are uniquely determined. The existence of the solutions follows from the fact that we can construct a homogeneous space G/H with its canonical affine connection of the second kind as we have shown before. We have thereby completed the proof of Theorem 18. 1.

20. Final remarks. First we shall complete the proof of b) $\rightarrow$ c) in Theorem 17. 1. By Theorem 18. 1, we see that the affine connection in question is locally representable as G/H with its canonical affine connection of the second kind. Since, in particular, $T = 0$, we see from the construction of the Lie algebra $\mathfrak{g} = \mathfrak{m} + \mathfrak{h}$ that $[\mathfrak{m}, \mathfrak{m}] \subset \mathfrak{h}$. This means that G/H is locally representable as a symmetric homogeneous space as we see from the proof of Theorem 15. 2. This concludes the proof of Theorem 17. 1.

To sum up, we have the following table which shows a close relationship between the simple types of affine connection and the groups spaces or homogeneous spaces.

a) $T = 0$, $R = 0$ (locally flat) $\rightleftarrows$ locally representable as an abelian group with its canonical affine connection.

b) $\Delta T = 0$, $R = 0 \rightleftarrows$ locally representable as a group with its $(-)$-connection.

c) $T = 0$, $\Delta R = 0$ (affine symmetric space) $\rightleftarrows$ locally representable as a symmetric homogeneous space.

d) $\Delta T = 0$, $\Delta R = 0$ (affine connection invariant by its parallelism) $\rightleftarrows$ locally representable as a reductive homogeneous space with its canonical affine connection of the second kind.

COLUMBIA UNIVERSITY.

BIBLIOGRAPHY.

[1] A. Borel and A. Lichnerowicz, "Espace riemanniens et hermitiens symétriques," *Comptes Rendus de l'Academie des Sciences (Paris)*, vol. 234 (1952), pp. 2332-2334.

[2] E. Cartan, "La géometrie des groupes de transformations," *Journal de Mathématiques pures et appliquées*, t. 6 (1927), pp. 1-119.

[3] E. Cartan, *La théorie des groupes finis et continus et l'analysis situs*, Memorial des Sciences Mathématiques, fasc. XLII (1930).

[4] ———, *Leçons sur la géométrie des espaces de Riemann*, Paris, Gauthier-Villars (1946).

[5] ———, "Sur une classe remarquable d'espaces de Riemann," *Bulletin de la Société Mathématique de France*, t. 54 (1926), pp. 214-264; *ibid.*, t. 55 (1927), pp. 114-134.

[6] S. S. Chern, *Lecture notes (Differential geometry)*, Department of Mathematics, University of Chicago.

[7] C. Chevalley, *Theory of Lie groups* I, Princeton University Press (1946).

[8] L. P. Eisenhart, *Continuous groups of transformations*, Princeton University Press (1933).

[9] K. Nomizu, "On the group of affine transformations of an affinely-connected manifold," to appear in *Proceedings of the American Mathematical Society*.

[10] S. Sasaki, "On a theorem concerning the homological structure end the holonomy groups of closed orientable symmetric spaces," *Proceedings of the Japanese Academy*, vol. 27, No. 27 (1951), pp. 81-85.

[11] J. H. C. Whitehead, "Affine spaces of paths which are symmetric about each point," *Mathematische Zeitschrift*, vol. 35 (1932), pp. 644-659.

[12] K. Yano, "On harmonic and Killing vector fields," *Annals of Mathematics* (2), vol. 55 (1952), pp. 38-45.

SOME THEOREMS CONCERNING SPIRALS IN THE PLANE.*

By B. J. Ball.

Throughout this paper the space under consideration will be the Euclidean plane.

The definitions given by R. L. Moore in [6] are essentially equivalent to the following, also due to Professor Moore. An arc t is said to *spiral down* on a point O provided there does not exist a finite sequence of rays [1] each starting from O such that the first has no point other than O in common with t, the last is a straight ray, and no two adjacent rays of the sequence have a point other than O in common. The arc t is said to *go n steps toward spiraling down on O* provided there does not exist a sequence of n rays satisfying the above conditions. An arc which spirals down on some point is called a *spiral* and a point set which contains no spiral is said to be *spiral free.*

In this paper an example is given of a compact, totally disconnected, closed point set such that every arc containing it spirals down on each of uncountably many points. In order that the compact, totally disconnected, closed point set M should be a subset of a spiral free arc it is sufficient that the projection of M onto some straight line be totally disconnected and it is necessary and sufficient that M be a subset of the sum of the elements of a continuous and equicontinuous collection G of mutually exclusive spiral free arcs such that G, with respect to its elements as points, is totally disconnected. A number of related results are also obtained.

Definitions. 1. A finite sequence F of simple discs [2] is called a *weak chain* provided that every two adjacent terms of F have only one point in common and every two nonadjacent terms of F are mutually exclusive. The terms of F are called its *links* and the first and last terms of F are called its *end links.* The sum of the links of F will be denoted by F^*, and in general, if G is a collection of point sets, G^* will denote the sum of the sets of the

* Received April 13, 1953; revised August 18, 1953.

[1] The term *ray* is here used to denote a closed topological image of a straight ray.

[2] A simple closed curve plus its interior is called a *simple disc.*

collection G. 2. The weak chain F is said to have *property* Y with respect to the simple disc K provided there are two points A and B of the boundary of K such that (1) there is only one link of F containing A and only one containing B and these are the end links of F and (2) $F^* - (A + B)$ is a subset of the interior of K. A point x of a link f of F will be called a *special point* of f with respect to F and K if x belongs either to the boundary of K or to a link of F different from f. 3. If K is a simple disc and n is a positive integer, then the weak chain F is said to have *property* X_n with respect to K provided that (1) F has property Y with respect to K, (2) F has only three links and (3) if O is a point of the center link of F and t is an arc which intersects both the boundary of K and the center link of F and does not go n steps toward spiraling down on O, then if f denotes either of the end links of F, $t \cdot f$ contains an arc which separates the special points of f from each other in f. It can be shown that if A and B are two points of the boundary of a simple disc K and n is a positive integer, then there exists a weak chain F such that F^* contains $A + B$ and F has property X_n with respect to K.

THEOREM 1. *If D is a domain, ϵ is a positive number and n is a positive integer, then there exist a compact, totally disconnected closed point set M and a collection G of mutually exclusive simple discs such that* (1) *$G^* + M$ is a closed subset of D,* (2) *each element of G has diameter less than ϵ and* (3) *if t is an arc containing M, there is an element g of G such that t goes n steps toward spiraling down on every point of g.*

Proof. Let K be a simple disc of diameter less than ϵ lying in D and let F be a weak chain having property X_n with respect to K. If f is an end link of F, there exists a weak chain c having property Y with respect to f such that c^* contains the special points of f and each link of c has diameter less than 1. Let C_1 denote a collection consisting of one such weak chain for each end link of F. If q is a link of an element of C_1, there exists a weak chain h having property X_n with respect to q such that h^* contains the special points of q. Let H_1 denote a collection consisting of one such weak chain for each simple disc which is a link of an element of C_1. For each end link f of an element of H_1 there exists a weak chain c having property Y with respect to f such that c^* contains the special points of f and each link of c has diameter less than $1/2$. Let C_2 denote a collection consisting of one such weak chain for each simple disc which is an end link of an element of H_1. If q is a link of a weak chain of the collection

C_2, there is a weak chain h having property X_n with respect to q such that h^* contains the special points of q. Let H_2 denote a collection consisting of one such weak chain for each simple disc which is a link of an element of C_2. This process may be continued. Hence there exist two sequences $H_0, H_1, H_2, \cdots$ and $C_1, C_2, C_3, \cdots$ such that H_0 is the collection whose only element is F and for each positive integer i, the following statements are true: (1) C_i and H_i are finite collections of weak chains, (2) every link of an element of C_i has diameter less than $1/i$, (3) if c is an element of C_i, there is an end link f of an element of H_{i-1} such that c has property Y with respect to f and c^* contains the special points of f, (4) if f is an end link of an element of H_{i-1}, there is exactly one element c of C_i such that c has property Y with respect to f and c^* contains the special points of f, (5) if h is an element of H_i, there is a link q of an element of C_i such that h has property X_n with respect to q and h^* contains the special points of q, (6) if q is a link of an element of C_i, there is exactly one element h of H_i such that h has property X_n with respect to q and h^* contains the special points of q.

For each non-negative integer i, let M_i denote the sum of all simple discs f such that f is an end link of an element of H_i. Then M_i is closed and contains M_{i+1}. Let M denote the common part of the point sets $M_0, M_1, M_2, \cdots$. Then M is closed and compact, and contains all the special points of the end links of the elements of $H_0, H_1, H_2, \cdots$. For each positive integer i, every component of M_i is a subset of the sum of two intersecting links of an element of C_i and consequently has diameter less than $2/i$. Hence M is totally disconnected. Let G denote the collection of all simple discs g such that for some non-negative integer i, g is the center link of an element of H_i. The elements of G are mutually exclusive. If i is a positive integer and g is the center link of an element of H_i, then every point of g is at a distance less than $1/i$ from some point of M. Hence $G^* + M$ is closed.

Suppose t is an arc containing M and there is no element g of G such that t goes n steps toward spiraling down on each point of g. There exists a ray r starting from a point O of the center link of F and not intersecting t such that no arc of r goes n steps toward spiraling down on O. Let f_0 denote one of the end links of F and let x and y denote its special points. Since F has property X_n with respect to K and r contains an arc intersecting both the boundary of K and the center link of F, r contains an arc which separates x from y in f_0.

Since r does not intersect M, there exists a positive integer i such that for no end link f of an element of H_i does r contain an arc which separates

the special points of f from each other in f. Let j denote the smallest such integer i. There exist a simple disc f, two points x and y of f, and an arc u of r such that f is one of the end links of an element of H_{j-1}, x and y are the special points of f, and u is a subset of f which separates x from y in f. Let c be the element of C_j which has property Y with respect to f. Since x and y belong to c^* and u separates x from y in f and does not contain any of the special points of the links of c, there exists a link q of c such that u contains an arc which separates the special points of q from each other in q. Let z and w be the special points of q and let v be a subarc of u which is irreducible with respect to the property of separating z from w in q. Let h denote the element of H_j which has property X_n with respect to q and let f_1 and f_2 denote the end links of h containing z and w, respectively. Let L denote the component of $q - v$ which contains z and let N denote the other component of $q - v$. Since the special points of f_1 belong to M and v does not intersect M and contains no arc which separates the special points of f_1 from each other in f_1, both of the special points of f_1 belong to L. Similarly, both special points of f_2 belong to N. Since both special points of f_1 belong to L and one of them belongs to the boundary of q and the other belongs to the center link of h, t contains an arc t' which is a subset of L and intersects both the boundary of q and the center link of h. Since h has property X_n with respect to q and there is a point of the center link of h such that t' does not go n steps toward spiraling down on it, t' contains an arc which separates the special points of f_2 from each other in f_2. Since t' is a subset of L and L does not contain either of the special points of f_2, $L \cdot f_2$ separates the special points of f_2 from each other in f_2 and hence so does $v \cdot f_2$. This involves a contradiction. Hence if t is an arc containing M there is an element g of G such that t goes n steps toward spiraling down on every point of g.

COROLLARY 1.1. *If A and B are two points, then for every positive integer n there exists a compact, totally disconnected, closed point set M_n not containing A or B such that every arc from A to B which does not intersect M_n goes at least n steps toward spiraling down on some point.*

Proof. The argument for Theorem 1 can be used almost without modification to establish that if P and Q are two points of the boundary of a simple disc K, then for every positive integer n there exists a totally disconnected, closed subset H_n of K such that every arc lying in $K - H_n$ and separating P from Q in K goes at least n steps toward spiraling down on some point. For any two points A and B there exist simple discs K_1 and K_2

with only two points, P and Q, in common such that $K_1 + K_2$ separates A from B. If H_{n1} and H_{n2} are compact, totally disconnected, closed subsets of K_1 and K_2, respectively such that for $i = 1$ or 2, every arc lying in $K_i - H_{ni}$ and separating P from Q in K_i goes n steps toward spiraling down on some point, then every arc from A to B which does not intersect $H_{n1} + H_{n2}$ goes n steps toward spiraling down on some point.

THEOREM 2. *There exists a compact, totally disconnected, closed point set such that every arc containing it spirals down on each of uncountably many points.*

Proof. It follows from Theorem 1 that there exist a compact, totally disconnected, closed point set M_1 and a collection H_1 of mutually exclusive simple discs such that (1) $M_1 + H_1^*$ is closed and compact, (2) each element of H_1 has diameter less than 1, (3) if t is an arc containing M_1, then there are two elements of H_1 such that t goes 1 step toward spiraling down on every point of their sum. If g is an element of H_1, there exist a closed, bounded, totally disconnected point set M_g and a collection H_g of mutually exclusive simple discs such that (1) $M_g + H_g^*$ is a closed subset of g, (2) each element of H_g has diameter less than $1/2$, (3) if t is an arc containing M_g, there are two elements of H_g such that t goes 2 steps toward spiraling down on every point of their sum. Hence there exist a point set M_2 and a collection H_2 of mutually exclusive simple discs each of diameter less than $1/2$ such that (1) $M_1 + M_2$ is closed and totally disconnected (2) H_1^* contains $M_2 + H_2^*$ and M_2 intersects every element of H_1 (3) if g is an element of H_1 and t is an arc containing $g \cdot M_2$, then there are two elements of H_2 lying in g such that t goes 2 steps toward spiraling down on every point of their sum.

Thus there exist two sequences $M_1, M_2, M_3, \cdots$ and $H_1, H_2, H_3, \cdots$ such that if n is a positive integer, then (1) M_n is a point set and H_n is a collection of mutually exclusive simple discs each of diameter less than $1/n$, (2) $M_1 + M_2 + M_3 + \cdots + M_n$ is closed and totally disconnected and $M_1 + M_2 + M_3 + \cdots + M_n + H_n^*$ is closed and compact, (3) H_n^* contains $M_{n+1} + H_{n+1}^*$ and M_{n+1} intersects every element of H_n, (4) if g is an element of H_n and t is an arc containing $g \cdot M_{n+1}$, then there are two elements of H_{n+1} lying in g such that t goes n steps toward spiraling down on every point of their sum.

For each positive integer n, let $M_n' = M_1 + M_2 + \cdots + M_n + H_n^*$. Then M_n' is closed and compact, each component of M_n' has diameter less

than $1/n$, and M_n' contains M_{n+1}'. If M denotes the common part of the point sets $M_1', M_2', M_3', \cdots$, then M is closed, compact and totally disconnected. Suppose t is an arc containing M. Since t contains M_1, there exist two simple discs g_1 and g_2 of H_1 such that t goes 1 step toward spiraling down on every point of their sum. Since t contains $M_2 \cdot g_1$ and $M_2 \cdot g_2$, there exist four simple discs g_{11}, g_{12}, g_{21} and g_{22} of H_2 such that g_{11} and g_{12} lie in in g_1, g_{21} and g_{22} lie in g_2, and t goes 2 steps toward spiraling down on every point of their sum. This process may be continued. Hence there exists a sequence $G_1, G_2, G_3, \cdots$ such that if n is a positive integer, then (1) G_n is a collection of 2^n mutually exclusive simple discs, (2) each element of G_{n+1} is a subset of some element of G_n and each element of G_n contains two elements of G_{n+1} and (3) t goes n steps toward spiraling down on each point of G_n^*. Let K denote the common part of the point sets $G_1^*, G_2^*, G_3^*, \cdots$. Then K is an uncountable subset of M. If O is a point of K, then for every positive integer n, t goes n steps toward spiraling down on O. Hence t spirals down on every point of K.

THEOREM 3. *If the arc t spirals down on the point O, then O belongs to t and every subarc of t which contains O spirals down on O.*

THEOREM 4. *If two arcs have only the point O in common and one of them spirals down on O, so does the other.*

Definition. Suppose M is either a domain or a nondegenerate continuous curve. The reversibly continuous transformation T of M into some point set is said to be *spiral preserving* provided that if u is an arc lying in M and P is a point of M, then u spirals down on P if and only if $T(u)$ spirals down on $T(P)$.

THEOREM 5. *A reversibly continuous transformation of a domain D into a domain is spiral preserving provided every point of D belongs to a spiral free arc which is transformed into a spiral free arc.*

Proof. Suppose u is an arc lying in D and P is a point of D such that u does not spiral down on P. There exists a sequence $u_1, u_2, \cdots, u_n$ of arcs lying in D such that $u_1 \cdot u = P, u_2 \cdot u_1 = P, \cdots, u_n \cdot u_{n-1} = P$ and u_n is a straight line interval. Suppose $T(u)$ spirals down on $T(P)$. Since $T(u)$ and $T(u_1)$ have only $T(P)$ in common and $T(u)$ spirals down on $T(P)$ it follows from Theorem 4 that $T(u_1)$ spirals down on $T(P)$. Since $T(u_1)$ and $T(u_2)$ have only $T(P)$ in common, $T(u_2)$ spirals down on $T(P)$. Hence

$T(u_n)$ spirals down on $T(P)$. Let g be a spiral free arc lying in D and containing P such that $T(g)$ is spiral free. There exists a sequence $g_1, g_2, \cdots, g_m$ of arcs lying in D such that $g_1 \cdot g = P, g_2 \cdot g_1 = P, \cdots, g_m \cdot g_{m-1} = P$ and g_m is a straight line interval. Since $T(g)$ and $T(g_1)$ have only $T(P)$ in common and $T(g)$ does not spiral down on $T(P)$, it follows from Theorem 4 that $T(g_1)$ does not spiral down on $T(P)$. Since $T(g_1)$ and $T(g_2)$ have only $T(P)$ in common, $T(g_2)$ does not spiral down on $T(P)$. By induction, $T(g_m)$ does not spiral down on $T(P)$. Since $T(u_n)$ spirals down on $T(P)$ and $T(g_m)$ does not, $T(u_n)$ and $T(g_m)$ have a point different from $T(P)$ in common. Consequently, since u_n and g_m are straight line intervals, there exists an interval v which contains P and is a subset of both u_n and g_m. Since $T(v)$ is a subarc of $T(u_n)$ containing $T(P)$ and $T(u_n)$ spirals down on $T(P)$, by Theorem 3, $T(v)$ spirals down on $T(P)$. It follows that $T(g_m)$ spirals down on $T(P)$. This is a contradiction, and hence $T(u)$ does not spiral down on $T(P)$.

It follows from the above argument that if v is an arc lying in $T(D)$ and Q is a point of $T(D)$ such that v does not spiral down on P, then $T^{-1}(v)$ does not spiral down on $T^{-1}(Q)$. Hence if u is an arc lying in D and P is a point of D such that u spirals down on P, then $T(u)$ spirals down on $T(P)$. Hence the transformation T is spiral preserving.

COROLLARY 5. 1. *Every conformal mapping of one domain onto another is spiral preserving.*

This is an immediate consequence of Theorem 5 and the fact that no conformal mapping transforms a straight line interval into an arc which spirals down on a point.

Definitions. A collection G of points sets is said to be *continuous* provided it is true that if the element g of G contains a point of the sequential limiting set of a sequence of elements of G, then g is the sequential limiting set of that sequence. A collection G of arcs is said to be *equicontinuous* provided that for every positive number ϵ there exists a positive number δ such that if x and y are two points of an arc g of G at a distance apart less than δ, then the diameter of the interval xy of g is less than ϵ.

THEOREM 6. *If A, B, D, and C are four points lying in that order on the simple closed curve J and no arc of J spirals down on any point, then there is a continuous and equicontinuous collection of G of mutually exclusive*

spiral free arcs filling up J plus its interior such that each arc of G is irreducible from AB to CD.

Proof. Let I denote the interior of J and let J' and I' denote a circle and its interior, respectively. There is a reversibly continuous transformation T of $I + J$ into $I' + J'$ which represents I conformally on I' ([2], pp. 81-86). (The use of this theorem was suggested by Professor E. J. McShane.) By Corollary 5. 1, T is spiral preserving over I. Since no arc of J spirals down on any point, T is spiral preserving over $I + J$. Let A', B', C', and D' denote, respectively, $T(A)$, $T(B)$, $T(C)$ and $T(D)$. It can be shown that there is a continuous and equicontinuous collection G of mutually exclusive spiral free arcs filling up $I' + J'$ such that each arc of G is irreducible from $A'B'$ to $C'D'$. The collection $T^{-1}(G)$ is a continuous and equicontinuous collection of mutually exclusive spiral free arcs filling up $I + J$ and each arc of $T^{-1}(G)$ is irreducible from AB to CD.

THEOREM 7. *Let J denote a rectangle with vertices A, B, D, C, such that the interval AB is vertical, A is above B, C is above D, and D is to the right of B. Let I denote the interior of J. Suppose G is a collection of mutually exclusive arcs such that* (1) *each arc of G is irreducible from AC to BD and is a subset of $J + I$, and* (2) *each pair of points of $J + I$ which lie on the same horizontal line are separated in $J + I$ by an element of G. Then there exists a continuous and equicontinuous collection G' of mutually exclusive arcs such that* (1) $G'^* = J + I$, (2) *every arc of G' is irreducible from AC to BD,* (3) *no horizontal line contains two points of any arc of G',* (4) *G is a subcollection of G'.*

LEMMA 7. 1. *Under the hypothesis of Theorem 7, G is equicontinuous.*

Proof. Suppose G is not equicontinuous. It follows that there exist a sequence $g_1, g_2, g_3, \cdots$ of arcs of G and three sequences $x_1, x_2, x_3, \cdots$, $y_1, y_2, y_3, \cdots$, and $z_1, z_2, z_3, \cdots$ such that (1) for each n, x_n, z_n, y_n lie in that order on g_n with x_n above z_n and z_n above y_n, (2) as $n \to \infty$, x_n and y_n converge to a point x and z_n converges to a point z different from x. Let l denote the horizontal line through x. Since for each n, x_n is above z_n and z_n is above y_n, it follows that z is on l. Since x and z are on the same horizontal line, there is an arc of G which separates x from z in $J + I$. This involves a contradiction. Hence G is equicontinuous.

Proof of Theorem 7. Suppose g is the sequential limiting set of a sequence α of arcs of G. Then g is a continuum lying in $J + I$ which inter-

sects both AC and BD. Suppose g' is an element of G which intersects g. If g is not a subset of g', there exist points x and y on the same horizontal line such that x belongs to g and y belongs to g'. Let g'' be an element of G which separates x from y in $J+I$ and let K_1 and K_2 be the components of $(J+I)-g''$ containing x and y respectively. Since K_1 is an open set relative to $J+I$ and contains a point of the sequential limiting set of α, and the boundary of K_1 relative to $J+I$ does not intersect more than one element of α, K_1 contains all but a finite number of the elements of α. But this is impossible since K_2 contains g' and g' intersects g. An almost identical argument shows that g cannot contain two points of any horizontal line. It follows that g is an arc irreducible from AC to BD and is either an element of G or does not intersect G^*.

Let G' denote the collection of all arcs g such that g either belongs to G or is the sequential limiting set of some sequence of arcs of G. It is easily see that the collection G' satisfies the desired conditions.

THEOREM 8. *Let A, B, C, D, J, and I be defined as in the hypothesis of Theorem 7 and suppose M is a closed, totally disconnected point set lying in $J+I$ such that the projection of M onto AB is totally disconnected. Then there exists a continuous and equicontinuous collection G of mutually exclusive arcs such that* (1) *$G^* = J+I$,* (2) *each arc of G is irreducible from AC to BD,* (3) *no horizontal line contains two points of any arc of G,* (4) *the common part of BD and the sum of the elements of G which intersect M is totally disconnected.*

LEMMA 8.1. *Under the hypothesis of Theorem 8, if h and k are two arcs lying in $J+I$, each irreducible from AC to BD and neither containing two points of any horizontal line, then there exists an arc t lying between h and k such that t does not intersect M, is irreducible from AC to BD and does not contain two points of any horizontal line.*

Proof. There exists a sequence $M_1, M_2, M_3, \cdots$ of points sets such that (1) for each positive integer n, M_n contains M_{n+1}, (2) for each n, M_n has only a finite number of components and if m is a component of M_n, then m is a simple disc bounded by an interval of AB, an interval of CD, and two horizontal intervals, (3) the common part of $M_1, M_2, M_3, \cdots$ is the set of all horizontal intervals with end points on AB and CD which intersect M. Suppose that for each positive integer n, there is a component m of M_n such that there is no vertical interval lying in $m - m \cdot M$ and separating $m \cdot h$

from $m \cdot k$ in m. It follows ([5], 56, Theorem 78) that there exists a sequence $m_1, m_2, m_3, \cdots$ such that for each positive integer n, (1) m_n is a component of M_n, (2) m_n contains m_{n+1}, (3) every vertical interval separating $m_n \cdot h$ from $m_n \cdot k$ in m_n intersects M. If l is the horizontal line containing the common part of $m_1, m_2, m_3, \cdots$, then every point of l between $l \cdot h$ and $l \cdot k$ belongs to M. But this is not possible since M is totally disconnected. Hence there exists an integer n_0 such that if m is a component of M_{n_0}, there is a vertical interval lying in m and not intersecting M and separating $m \cdot h$ from $m \cdot k$ in M. Hence there exists an arc t irreducible from AC to BD such that (1) t lies between h and k, (2) if m is a component of M_{n_0}, then $t \cdot m$ is a vertical interval not intersecting M and separating $m \cdot h$ from $m \cdot k$ in m, (3) if D is a component of $I - I \cdot M_0$, then $t \cdot \bar{D}$ is an arc which does not contain two points of any horizontal line. It follows from (2) and (3) that no horizontal line contains two points of t.

Proof of Theorem 8. Suppose H is a finite collection of mutually exclusive arcs each irreducible from AC to BD and lying in $J + I$ and no arc of H contains two points of any horizontal line. Then the *maximum width* of a component D of $(J + I) - H^*$ will mean the length of the longest horizontal straight line interval lying in $\bar{D}$. Let d denote the length of AC. With the aid of Lemma 8.1 it can be shown that there exists a sequence $G_1, G_2, G_3, \cdots$ such that for each n, (1) G_n is a finite collection of mutually exclusive arcs each lying in $(J + I) - M$ and irreducible from AC to BD, (2) no element of G_n contains two points of any horizontal line, (3) G_n is a subcollection of G_{n+1}, (4) if D is a component of $(J + I) - G_n^*$, then the maximum width of D is less than $(3/4)^n \cdot d$. Let $G = G_1 + G_2 + \cdots$.

Suppose x and y are two points of $J + I$ lying on the same horizontal line. There exists an integer n such that $(3/4)^n \cdot d < \rho(x, y)$ and hence some element of G_n separates x from y in $J + I$. Hence G satisfies the hypothesis of Theorem 7 and consequently there exists a continuous and equicontinuous collection G' of mutually exclusive arcs such that (1) $G'^* = J + I$, (2) every arc of G' is irreducible from AC to BD, (3) no horizontal line contains two points of any element of G', (4) G is a subcollection of G'. Since G^* does not intersect M and is dense in $J + I$, the common part of BD and the sum of the elements of G' which intersect M is totally disconnected.

THEOREM 9. *If M is a compact, totally disconnected, closed point set and the projection of M onto the y-axis is totally disconnected, then there exists a reversibly continuous, spiral preserving transformation T of the plane*

into itself which carries each horizontal line into itself and is such that the projection of $T(M)$ onto each of the coordinates axes is totally disconnected.

Proof. There exists a rectangle J with interior I and vertices A, B, D, C, such that M is a subset of I and AB is vertical, A is above B, C is above D and D is to the right of B. There exists a continuous and equicontinuous collection G of mutually exclusive arcs satisfying the conditions of Theorem 8 with respect to J, I, A, B, C, D, and M. There exists a reversibly continuous transformation T of the plane into itself which carries each arc of G into a vertical straight line interval and which carries each horizontal line into itself (See [4], Theorem 1). By Theorem 5, T is spiral preserving. It may be easily seen that the transformation T carries M into a point set whose projections onto the coordinate axes are totally disconnected.

THEOREM 10. *Suppose R is a rectangle with horizontal and vertical sides and G is a finite collection of mutually exclusive rectangular discs with horizontal and vertical edges such that G^* is a subset of the interior of R. Let AB and CD denote the vertical sides of R. There exists a finite collection H of mutually exclusive arcs such that* (1) *each arc of H has one end point on $AB - (A + B)$ and the other on $CD - (C + D)$ and lies except for its end points wholly in the interior of R,* (2) *no vertical line contains two points of any arc of H,* (3) *if g is an element of G, there exist two arcs h_g and k_g of H such that h_g contains the upper edge, and k_g the lower edge, of g and* (4) *no element of H intersects two elements of G.*

THEOREM 11. *If the totally disconnected point set M is the common part of a sequence $M_1, M_2, M_3, \cdots$ of point sets such that for each n, M_n is the sum of a finite number of mutually exclusive rectangular discs with horizontal and vertical edges, then there exists a spiral free arc containing M.*

Proof. Under the above hypothesis, M is the common part of a sequence $M_1', M_2', M_3', \cdots$ of point sets such that for each n, (1) M_n' is the sum of a finite number of mutually exclusive rectangular discs, (2) every component of M_n' intersects M and (3) every component of M_{n+1}' is a subset of the interior of some component of M_n'.

Let R denote a rectangle with horizontal and vertical sides whose interior, I, contains M_1'. Let A, B, D, C denote vertices of R such that A is above B, C is above D, and D is to the right of B. Let M_0' denote $R + I$ and let G_1 denote the collection of all components of M_1'. There exists a finite collection H_1 of arcs which satisfies the conditions of Theorem 10 with

respect to R and G_1. If m is a component of M_1', there exists an arc t which is the sum of the left edge of m and two arcs t_1 and t_2 each of which is irreducible from m to AB and is a subset of some arc of H_1. Hence there exists an arc u_1 from A to B which is a subset of $AB + H_1^* + M_1'$ and which contains the left edge of each component of M_1'. Let K_1 denote the point set $u_1 + M_1'$.

Let $m_1, m_2, \cdots, m_n$ denote the components of M_1'. For each positive integer i not greater than n, let R_i denote the boundary of m_i and let x_iy_i denote the left edge of m_i. Let G_{2i} denote the collection of all components of $M_2' \cdot m_i$. There exists a finite collection H_{2i} of arcs which satisfies the conditions of Theorem 10 with respect to R_i and G_{2i}. By an argument similar to that given above, it can be shown that there exists an arc u_{2i} from x_i to y_i which is a subset of $x_iy_i + H_{2i} + m_i \cdot M_2'$ and which contains the left edge of each component of $m_i \cdot M_2'$. Let K_2 denote the point set

$$(K_1 - M_1') + (u_{21} + u_{22} + \cdots + u_{2n}) + M_2'.$$

This process may be continued.

Hence there exists a sequence $K_1, K_2, K_3, \cdots$ of continua such that for each positive integer n, (1) K_n contains K_{n+1}, (2) K_n is the sum of M_n' and an arc u_n from A to B which contains the left edge of each component of M_n', (3) if P is a point of K_n which is not in the interior of any component of M_n', then there is a ray r starting from P such that no vertical line contains two points of r and such that $r - P$ does not intersect K_n.

Let K denote the common part of the point sets $K_1, K_2, K_3, \cdots$. Then K is a compact continuum containing $M + A + B$. It may be shown that K contains no continuum of condensation and hence is a continuous curve. If P is a point of $K - (M + A + B)$, then for some n, P does not belong to M_n'. It follows that P separates A from B in M_n' and therefore separates A from B in K. Since for each n, every component of M_n' separates A from B in K_n and K is a continuous curve, it follows that every point of M separates A from B in K. Hence K is an arc from A to B.

If P is a point of K, there is a ray r starting from P such that no vertical line contains two points of r and such that $r - P$ does not intersect K. Consequently K does not spiral down on any point.

THEOREM 12. *If G is a continuous and equicontinuous collection of mutually exclusive, spiral free arcs such that G^* is closed and compact and does not separate the plane, then there exist a rectangle $ABDC$ and a reversibly continuous, spiral preserving transformation of the plane into itself which*

carries each arc of G into a straight line interval which has its end points on AB and CD and is parallel to BD.

Proof. There exist a simple closed curve J and four points A, B, D, C lying in that order on J such that every arc G is irreducible from AB to CD and lies except for its end points wholly in I, the interior of J ([1], Theorem 12). With the aid of Theorem 6, the arguments for Theorems 11 and 13 of [1] can easily be modified to show that there exists a continuous and equicontinuous collection G_1 of mutually exclusive arcs such that (1) $G_1^* = J + I$, (2) each arc of G_1 is irreducible from AB to CD, (3) G is a subcollection of G_1, (4) no arc of G_1 spirals down on any point of I.

There exist two mutually exclusive arcs u_1 and u_2 such that (1) A and B are the end points of u_1, u_1 contains $G^* \cdot AB$, and $u_1 - (A + B + G^* \cdot AB)$ is a subset of I, (2) C and D are the end points of u_2, u_2 contains $G^* \cdot CD$, and $u_2 - (C + D + G^* \cdot CD)$ is a subset of I, (3) no arc of G_1 contains two points of u_1 or of u_2. Let G_2 denote the collection of all arcs g such that g is a subarc of some arc of G_1, and is irreducible from u_1 to u_2. Then G_2 is a continuous and equicontinuous collection of mutually exclusive spiral free arcs, G_2^* is a simple disc, and G is a subcollection of G_2.

Let g_1 and g_2 denote the arcs of G_2 from A to C and from B to D, respectively. Let J_2 denote the boundary of G_2^*. Then $J_2 = u_1 + u_2 + g_1 + g_2$. Let J_1 denote a rectangle which has G_2^* in its interior. Let R denote the connected domain bounded by $J_1 + J_2$ and let R' denote a connected domain bounded by two concentric circles J_1' and J_2' with J_2' in the interior of J_1'. There exists a reversibly continuous transformation T' of $R + J_1 + J_2$ into $R' + J_1' + J_2'$ which is conformal over R ([2], pp. 72 and 86). Since no arc of J_1 or of J_2 spirals down on any point, it follows that T' is spiral preserving over $R + J_1 + J_2$. The vertices of J_1 may be labeled A', B', D', C' in such a way that the points $T'(A')$, $T'(B')$, $T'(D')$, $T'(C')$ lie in that order on J_1' and are in the same sense on J_1' that the points $T'(A)$, $T'(B)$, $T'(D)$, $T'(C)$ are on J_2'. Let A'' and B'' be points lying in the order $A'A''B''B'$ on $A'B'$ and let C'' and D'' be points lying in the order $C'C''D''D'$ on $C'D'$. There exist four mutually exclusive spiral free arcs each lying in R' except for their end points and such that one of them has end points $T'(A)$ and $T'(A'')$, one has end points $T'(B)$ and $T'(B'')$, one has end points $T'(C)$ and $T'(C'')$, and one has end points $T'(D)$ and $T'(D'')$. The inverses of these arcs are four mutually exclusive spiral free arcs AA'', BB'', CC'', and DD'', each lying in R except for its end points. These arcs divide

R into four simple domains. By applying Theorem 6 to each of these and adding together certain arcs, it can be shown that there exists a continuous and equicontinuous collection G' of mutually exclusive, spiral free arcs such that (1) G'^* is the sum of J_1 and its interior, (2) each arc of G' is irreducible from $A'B'$ to $C'D'$, (3) each arc of G_2 is a subarc of some arc of G'.

Suppose $A'B'$ is vertical. There exists a reversibly continuous transformation T of the plane into itself which carries each arc of G' into a horizontal straight line interval with end points on $A'B'$ and $C'D'$ and which does not change the x-coordinate of any point which is not in the interior of J_1 (See [4], Theorem 1). It follows from Theorem 5 that T is spiral preserving.

THEOREM 13. *In order that the compact, totally disconnected, closed point set M should be a subset of a spiral free arc it is necessary and sufficient that M be a subset of the sum of the elements of a continuous and equicontinuous collection G of mutually exclusive spiral free arcs such that G with respect to its elements as points is totally disconnected.*

Proof. The conditions are clearly necessary. The following argument shows that they are also sufficient.

With the help of the fact that every sequence of point sets in the plane which has limiting set contains a subsequence which has a sequential limiting set ([3], Theorem 5), it may be shown that if G' is the set of all elements of G which intersect M, then G'^* is closed and compact. Since G with respect to its elements as points is totally disconnected, every compoent of G^* is an arc of G. Hence G'^* does not separate the plane.

By Theorem 12, there exist a rectangle $ABDC$ and a reversibly continuous transformation T_1 of the plane into itself which carries each arc of G into a straight line interval which has its end points on AB and CD and is parallel to BD. Clearly, the projection of $T_1(M)$ onto AB is totally disconnected. Suppose AB is vertical. Since the projection of $T_1(M)$ onto AB is totally disconnected, it follows from Theorem 9 that there exists a reversibly continuous spiral preserving transformation T_2 of the plane into itself which carries each horizontal line into itself and which carries $T_1(M)$ into a point set whose projection onto each of the coordinate axes is totally disconnected. Let T denote the transformation T_2T_1 and let N denote the point set $T(M)$.

Since the projection of N onto each of the coordinate axes is totally disconnected, if q is a component of N and ϵ is a positive number, then there exists a rectangle R_q with horizontal and vertical sides which encloses q and

does not intersect N and is such that every point of it is at a distance less than ϵ from q. From this it may be shown that N satisfies condition (2) of the hypothesis of Theorem 11.

Hence there exists a spiral free arc K containing N. Since $N = T(M)$ and the transformation T is spiral preserving, $T^{-1}(K)$ is a spiral free arc containing M.

UNIVERSITY OF VIRGINIA.

REFERENCES.

[1] B. J. Ball, "Continuous and equicontinuous collections of arcs," *Duke Mathematical Journal*, vol. 19 (1952), pp. 423-433.

[2] C. Caratheodory, "Conformal representation," *Cambridge Tracts in Mathematics and Mathematical Physics*, No. 28, London, 1932.

[3] R. G. Lubben, "Concerning limiting sets in abstract spaces," *Transactions of the American Mathematical Society*, vol. 30 (1928), pp. 668-685.

[4] R. L. Moore, "Concerning certain equicontinuous systems of curves," *Transactions of the American Mathematical Society*, vol. 22 (1921), pp. 41-55.

[5] ———, "Foundations of point set theory," *American Mathematical Society Colloquium Publications*, vol. XIII, New York, 1932.

[6] ———, "Spirals in the plane," *Proceedings of the National Academy of Sciences*, vol. 39 (1953), pp. 207-213.

ON PARABOLIC CURVES ON SURFACES.*

By PHILIP HARTMAN and AUREL WINTNER.

Let S be a piece of a surface (of class C^2, at least) in an (x, y, z)-space, and let A be a parabolic arc on S, that is, an arc consisting of parabolic points, points at which the Gaussian curvature K of S vanishes. According to a consideration of P. Franklin [2], pp. 254-256, such an A (on an S of class C^3, say) is a line of curvature of S if the following three conditions are satisfied: (a) no point of A is "singular," and (b) no point of A is a flat point of S (that is, a point at which the matrix of the second fundamental form of S is the zero matrix), finally (c) at no point of S does a normal section of S possess a flex point.

It is not clear what Franklin means by "singular" in assumption (a). In this regard, it was shown in [3], Theorem (*), pp. 610-613, that if (a) is interpreted as meaning (a*) grad $K \neq 0$ along A (which makes (b) superfluous), while (c) is retained, then Franklin's assertion is certainly true but its truth is of a drastic nature; in fact, (a*) and (c) imply that A is a plane curve and, what is more, the tangent plane of S does not vary along A (so that A is an asymptotic line, as well as a line of curvature, of S). In what follows, this result of [3] will be paralleled by a similar but deeper fact, as follows:

(**) Let S: $z = z(x, y)$ be a surface of class C^2 defined over an open (x, y)-domain D. Suppose that there exists in D a Jordan arc of class C^1,

$$D^0\colon\ x = x(t),\ y = y(t)\,; \qquad -1 < t < 1, \tag{1}$$

which divides D into two domains D^+, D^- on which the respective inequalities $K > 0$, $K < 0$ hold for the Gaussian curvature $K = K(x, y)$ of S (so that, in particular,

$$A\colon\ (x, y, z) = (x(t), y(t), z(x(t), y(t)))\,; \qquad -1 < t < 1 \tag{1 bis}$$

is a parabolic arc of class C^1 on S). Suppose further that, if P is any point of A, the tangent plane $T = T(P)$ of S at P is a supporting plane of S (near P). Then A is a plane curve and, what is more, T is independent of P.

---

* Received February 28, 1953.

The assertion of this theorem (**) is the same as that of theorem (*) of [3], quoted above. But S is required to be of class C^3 in (*), and just of class C^2 in (**). Moreover, the assumption grad $K \neq 0$ of (*) is reduced in (**) to the much weaker assumption which requires merely the division (by D^0) of D into D^+ and D^-. Finally, assumption (c) of (*) is replaced by the last assumption of (**), requiring that S be supported by $T(P)$ at every point P of A. It is understood that (even though P is a cluster point of points of negative curvature) the tangent plane $T(P)$ is said to support S at P if S is contained in either of the closed half-spaces having the boundary $T(P)$ (at least if S is replaced by a sufficiently small neighborhood of P on S).

A theorem of A. D. Alexandroff (Theorem 1 in [1]) is concerned with parabolic arcs on certain *closed* surfaces and is a "theorem in the large," a fact underlying the proof given by Alexandroff. In what follows, his proof will be modified so as to become adjusted to the situation presented by the "local theorem" (**) above. Thus the proofs of the facts corresponding to those in Sections 5-6 of [1], p. 72, will have to be quite different from the arguments used (*loc. cit.*) by Alexandroff. On the other hand, the end of the proof of (**) will be identical with the proof in Section 7 of [1], pp. 72-73.

Remark. It is worth emphasizing that (**) is *not* a theorem on a piece of surface S with the properties that $K > 0$ on S and that a part of the boundary is a (smooth) arc A on which $K = 0$. For example, let $S\colon X = X(u, v)$, where $X = (x, y, z)$, be a surface of class C^2 defined on the partly closed set $(u \geqq 0, u^2 + v^2 < r^2)$ in such a way that $K = K(u, v) \geqq 0$ according as $u \geqq 0$ and that the tangent plane at a point $(u, v) = (0, v)$, where $v^2 < r^2$, supports S. It does not follow from (**), nor is it true, that (i) $A\colon X = X(0, v)$ is a plane curve or that (ii) if A is a plane curve, the tangent plane does not vary along it. This will be shown by two examples.

Ad (i). Consider the surface $S\colon z = x^2 + (x + \frac{1}{2}y^2)^4$, where (x, y) is in a vicinity D of a point $(x_0, y_0) \neq (0, 0)$ of the parabola $D_0\colon f(x, y) = 0$, where $f = x + \frac{1}{2}y^2$. Since $z_{xx} = 2 + 12f^2$, $z_{xy} = 12yf^2$, $z_{yy} = 4f^3 + 12y^2f^2$, the curvature is

$$K = K(x, y) = 4f^2(2x + 7y^2 + 12f^3)/(1 + z_x^2 + z_y^2)^2.$$

Hence, if D is sufficiently small, $K \geqq 0$ according as (x, y) is not or is on D_0. The arc $A\colon (-\frac{1}{2}y^2, y, y^4/4)$, for y near y_0, is not a plane curve, although S is convex, hence every tangent plane is a supporting plane.

In view of Franklin's statement quoted above, it is worth pointing out that A is not a line of curvature. In fact, $z_y = 0$, $z_{xx} = 2$, $z_{xy} = 0$ on A, so that the directions of principal curvature at a point of A are determined by $dx = 0$ and $dy = 0$ and, therefore, A is not in a direction of principal curvature (except at the point $(x, y) = (0, 0)$, which is not in D).

Ad (ii). Consider the surface S: $z = x^2 + (x + y)^4$. Since

$$z_{xx} = 2 + 12f^2, \qquad z_{xy} = 12f^2, \qquad z_{yy} = 12f^2,$$

where $f = x + y$, the curvature K is $24f^2/(1 + z_x^2 + z_y^2)^2$. Hence $K \geqq 0$ according as (x, y) is not or is on the line D_0: $x + y = 0$. The arc A: $(x, -x, x^2)$ is a plane curve but is not in a direction of principal curvature at any point. In particular, the normal vector $(2x, 0, -1)/(1 + 4x^2)^{\frac{1}{2}}$ is not constant along A.

Proof of (**). Let the notations be so chosen that D is a neighborhood of the point $(x, y) = (0, 0)$ and that this point is on D^0. If $r > 0$ is so small that the circle $B = B(r)$: $x^2 + y^2 < r^2$ is in D, let $E = E(r)$ be a neighborhood of $(0, 0)$ contained in B and having the following properties: E has a decomposition similar to that of D, that is, $E = E^+ + E^0 + E^-$, where $E^\pm$ are (non-empty) subsets of $D^\pm$, respectively, and E^0 is a subset of D^0 and consists of an (open) subarc of D^0. Thus $z = z(x, y)$ is defined on the closure of E, and the segment joining any pair of points in E is in D.

It can be supposed that the point $(x, y) = (0, 0)$ of D^0 corresponds to $t = 0$, that

$$z(0, 0) = z_x(0, 0) = z_y(0, 0) = 0, \tag{2}$$

and, since the tangent plane $z = 0$ at $(0, 0, 0)$ supports S, that

$$z(x, y) \leqq 0 \text{ on } D. \tag{3}$$

The normal vector $N = N(x, y) = (-z_x, -z_y, 1)/1 + z_x^2 + z_y^2)^{\frac{1}{2}}$ is of class C^1, since z is of class C^2. Hence, the assumption that D^0 is of class C^1 implies that $N = N(t) = N(x(t), y(t))$ is also of class C^1.

Suppose that the assertion of (**) is false, i. e., that $dN/dt \not\equiv 0$. It can be assumed that

$$dN/dt \neq 0 \tag{4}$$

at $t = 0$ and hence along $E^0 = E^0(r)$, if r is sufficiently small. According to (4), no point of E^0 is a flat point of S.

It will be shown that, if $r(> 0)$ is sufficiently small, and if (x, y),

$(x(t), y(t))$ are points of E^+, E^0, respectively, then $N(x, y) \neq N(t)$. It will be clear from the proof that it is sufficient to prove $N(x, y) \neq N(0, 0)$ for (x, y) in E^+.

Suppose, if possible, that $N(x_0, y_0) = N(0, 0) = (0, 0, 1)$ holds at a point (x_0, y_0) of E^+. Since $K(x_0, y_0) > 0$, the function z has a strict extremum at (x_0, y_0). This extremum is a maximum. In order to see this, it is sufficient to note that, since $K(0, 0) = 0$, the point $(0, 0)$ is not a flat point and (3) holds, it can be supposed, after a rotation of the (x, y)-plane, that $z(x, y) = -\frac{1}{2}ax^2 + o(x^2 + y^2)$ as $(x, y) \to (0, 0)$, where $a > 0$. Hence $z_{xx}(0, 0) = -a^2 < 0$, and so $z_{xx} < 0$ on E if r is sufficiently small.

After another rotation of the (x, y)-plane, it can be assumed that $x_0 > 0$ and $y_0 = 0$. The function $z(x, y)$ is defined on the segment $y = 0, 0 \leqq x \leqq x_0$, which is in D. It follows that $z(x, 0)$ has a maximum at $x = 0$ and at $x = x_0$. Consequently, the curvature $K = K(x, y)$ cannot satisfy $K(x, 0) > 0$ for $0 < x \leqq x_0$, for otherwise the curve $z = z(x, 0)$ in the plane $y = 0$ would be concave downwards. Let x_1 be the greatest value of $x (< x_0)$ at which $K(x, 0) = 0$. Thus $K(x, 0) > 0$ for $x_1 < x \leqq x_0$, and $(x_1, 0, z(x_1, 0))$ is a point of D^0. Clearly, $z(x_1, 0) < z(x_0, 0) \leqq 0$, $z_x(x_1, 0) > 0$, since the plane curve $z = z(x, 0)$ is concave downwards for $x_1 \leqq x \leqq x_0$. The tangent line $y = 0$, $L(x, z) = 0$, where $L(x, z) = z - z(x_1, 0) - (x - x_1)z_x(x_1, 0)$, separates $(x, z) = (0, 0)$ from $(x, z) = (x_0, z(x_0, 0))$ in the (x, z)-plane. In fact, $L(x_0, z(x_0, 0)) = z(x_0, 0) - z(x_1, 0) - (x_0 - x_1)z_x(x_1, 0) < 0$, by the concavity of the curve $z = z(x, 0)$, while $L(0, 0) = -z(x_1, 0) + x_1 z_x(x_1, 0) > 0$. Since the tangent plane to S at the point $(x, y) = (x_1, 0)$ contains the line $L(x, z) = 0$, $y = 0$, the assumption of (**), according to which the tangent plane of S at the point $(x_1, 0, z(x_1, 0))$ of D^0 supports S, cannot hold unless that tangent plane is $y = 0$. But the latter case is impossible if $r (> 0)$ is sufficiently small, since $N(0, 0) = (0, 0, 1)$. This contradiction proves that $N(x_0, y_0) \neq N(0, 0)$.

Since the proof of the last inequality depended only on the continuity of N and of the second partial derivatives of z at $(x, y) = (0, 0)$, and since the continuity is uniform on $x^2 + y^2 \leqq r^2$, it follows that $N(x, y) \neq N(x(t), y(t))$ if $r > 0$ is sufficiently small and if (x, y), $(x(t), y(t))$ are points of E^+, E^0, respectively.

Let $r > 0$ be fixed and so small that $N(x, y) \neq N(x(t), y(t))$ if $(x, y) \neq (x(t), y(t))$ is on the *closure* of E^+ and $(x(t), y(t))$ is on E^0; cf. (4). Let

$$E^1\colon\ x = x^1(t), y = y^1(t)\,; \qquad -1 < t < 1,$$

be an arc of class C^1 in E such that $(x^1(0), y^1(0)) = (0, 0)$, that $(x^1(t), x^2(t))$ is in $E^{\pm}$ according as $t \gtrless 0$, that E^1 crosses (does not touch) E^0 at $(x, y) = (0, 0)$, and that $dN(x^1(t), y^1(t))/dt \neq dN(x(t), y(t))/dt$ (as to the last condition, note that a directional derivative of N at $(0, 0)$ can be 0 only in the direction of principal curvature 0).

If $t > 0$ is sufficiently small, then no point (x^+, y^+) of E^+ can satisfy $N(x^+, y^+) = N(x^1(-t), y^1(-t))$. In order to prove this, suppose, if possible, that the last relation is satisfied by a sequence of values $t = t_1, t_2, \cdots$ satisfying $t_n \to +0$ as $n \to \infty$, where $(x^+, y^+) = (x_n, y_n)$ is in E^+. Then $N(x_n, y_n) \to N(0, 0)$ as $n \to \infty$, hence $(x_n, y_n) \to (0, 0)$ by virtue of the above choice of r and $E(r)$. Let $(x^1(t_n), y^1(t_n))$ and (x_n, y_n) be joined by a smooth Jordan arc J_n contained in E^+. Then the normal image of J_n is a locally smooth (possibly self-intersecting) arc which has the same end points as does the normal image of that subarc of E^1 on which $-t_n \leqq t \leqq t_n$. When n is sufficiently large, these end points are on "opposite sides" of the arc which is the normal image of E^0. On the other hand, the normal image of J_n has no point in common with the normal image of E^0. This leads to a contradiction, since it can be assumed that the arc-length of J_n and, hence, of its normal image, tends to 0 as $n \to \infty$.

What remains to be proved in order to complete the proof of (**) depends on the conclusions made by A. D. Alexandroff in Section 7, pp. 72-73 of his paper [1]; so that the balance of the proof of (**) will now be given only for the sake of completeness.

The plane $T = T_t$ tangent to S at a point $(x, y) = (x^1(t), y^1(t))$ of E^- must intersect S, since $K(x^1(t), y^1(t)) < 0$; cf., e. g., [3], pp. 617-619. If $-t < 0$ is sufficiently small, T_t contains no points (x, y, z) of S, where (x, y) is on the boundary of E. For otherwise T_0, the tangent plane at the point $(0, 0, 0)$ of S, would contain such a point (x_0, y_0). In view of (3), this would imply that $N(x_0, y_0) = N(0, 0)$ and that $K(x_0, y_0) \geqq 0$. But this is impossible in view of the above choice of r and $E = E(r)$.

Since T_t contains no points of (x, y, z) of S at which (x, y) is on the boundary of E, there is a (connected) domain S_t on S which lies on one side of the plane T_t and is such that its boundary points are on T_t (the projection of S_t on the (x, y)-plane is in E). Furthermore, S_t has a supporting plane parallel to T_t, and this supporting plane touches S_t at some point (x_0, y_0, z_0). Clearly, $K(x_0, y_0) \geqq 0$. But $K(x_0, y_0) = 0$ is impossible, since the normal images of the arcs E^0 and E^1 have only the point $N = (0, 0, 1)$ in common, and $K(x_0, y_0) > 0$ is impossible, since $N(x_0, y_0)$

$\neq N(x^1(-t), y^1(-t))$ for (x_0, y_0) in E^+ and $t < 0$ sufficiently small.

Since this contradiction resulted from the assumption (4), it follows that $dN/dt \equiv 0$, which proves (**).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. Alexandroff, "Über eine Klasse geschlossener Flächen" (in Russian, with a German abstract), *Recueil Mathématique*, vol. 46 (1938), pp. 69-77.

[2] P. Franklin, "Regions of positive and negative curvature on closed surfaces," *Journal of Mathematics and Physics*, vol. 13 (1934), pp. 253-260.

[3] P. Hartman and A. Wintner, "On geodesic torsions and parabolic and asymptotic curves," *American Journal of Mathematics*, vol. 74 (1952), pp. 607-625.

ON CONTINUOUS AREA-PRESERVING AND LEGENDRE TRANSFORMATIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

This note is concerned with minimal assumptions of differentiability for some classical theorems dealing with continuous, one-to-one mappings

$$T_1\colon\ u = u(x, y),\ v = (x, y) \tag{1}$$

which are of the Legendre or "semi-Legendre" type.

In what follows, all domains (i. e., open sets) are simply connected and, in fact, "small" vicinities of a point.

1. Area-preserving transformations. The first theorem to be proved is the following:

(I) *Let* $u = u(x, y)$, $v = v(x, y)$ *be continuous functions on some* (x, y)*-domain. Necessary and sufficient in order that the transformation* $T_1\colon (x, y) \to (u, v)$, *defined by* (1), *has the properties*

(i) T_1 *is one-to-one;*

(ii) $u(x, y)$ *is an increasing function of* y *for fixed* x*;*

(iii) T_1 *is orientation-preserving;*

(iv) T_1 *is area-preserving,*

is the existence of a function $f = f(x, u)$ *having the properties*

(α) f *has continuous partial derivatives* f_x, f_u*;*

(β) $f_x(x, u)$ *is an increasing function of* u *for fixed* x*;*

(γ) $f_u(x, u)$ *is an increasing function of* x *for fixed* u*;*

(δ) T_1 *is representable as the product* $T_1 = T_3 T_2^{-1}$, *where* T_2 *and* T_3 *are the respective (one-to-one) semi-Legendre transformations* $(x, u) \to (x, y)$ *and* $(x, u) \to (u, v)$ *defined by*

$$T_2\colon\ x = x,\ y = f_x(x, u) \tag{2}$$

and

$$T_3\colon\ u = u,\ v = -f_u(x, u). \tag{3}$$

* Received June 26, 1953.

This theorem is classical ([1]; cf. [4]) if it is assumed that u and v are of class C^1 (and, correspondingly, f is of class C^2) and that (ii) is replaced by $u_y > 0$ and (β), (γ) by $f_{ux} > 0$.

If a function $f(x, u)$ with the properties (α), (β), (γ) is given, then (2), (3) and the relation $T_1 = T_3 T_2^{-1}$ determine a unique transformation (1) with the properties (i)-(iv). Conversely, if (1) is given with the properties (i)-(iv) (on a simply connected (x, y)-domain), then an f satisfying (α)-(δ) is uniquely determined up to an additive constant.

Proof of (I). Let (1) define a continuous transformation T_1 with the properties (i)-(iv) and let the inverse of T_1 be denoted by

$$T_4\colon\ x = x(u, v),\ y = y(u, v). \tag{4}$$

Since $u(x, y)$ in (1) is an increasing function of y for fixed x, the equation $u = u(x, y)$ has a solution $y = y(x; u)$ for y. Let D denote the domain of definition of $y(x; u)$ (so that D is simply connected). Clearly,

$$T_5\colon\ x = x,\ y = y(x; u) \tag{5}$$

defines a topological mapping $(x, u) \to (x, y)$.

If $y = y(x; u)$ is substituted into the second of the functions (1), there results a function $v = v(x, y(x; u)) = v(x; u)$ of (x, u) on D. The transformation

$$T_6\colon\ u = u,\ v = v(x; u) \tag{6}$$

is topological, since $T_6 = T_1 T_5$. This shows that the property (ii) of the topological mapping T_1 implies that the inverse $T_4 = T_1^{-1}$ has the following property:

(ii′) $x = x(u, v)$ is a strictly monotone function of v for fixed u.

Actually, the monotone function $v(x; u)$ of x is decreasing, since T_1 is orientation-preserving. For the fact that $y(x; u)$ is an increasing function of u shows that T_5 is orientation-preserving and hence $T_6 = T_1 T_5$ is also. But T_6 is orientation-preserving if and only if the monotone function $v(x; u)$ of x is decreasing.

Let R denote a rectangle $x_1 \leqq x \leqq x_2$, $u_1 \leqq u \leqq u_2$ in D. Let R_5 and R_6 denote the (x, y)- and (u, v)-images of R under T_5 and T_6, respectively. Thus R_5 is a domain bounded by the two arcs $y = y(x; u_1)$, $y = y(x; u_2)$, where $x_1 \leqq x \leqq x_2$, and two segments of the lines $x = x_1$, $x = x_2$, and R_6 is a similarly described domain. Hence, the areas of R_5 and R_6 are given by

the respective line integrals $\int y(x;u)\,dx$, $\int v(x;u)\,du$ around the boundary of R. Since R_6 is the image of R_5 under (1), the area-preserving character of T_1 means that

$$\int y(x;u)\,dx - v(x;u)\,du = 0, \tag{7}$$

where the line integral is taken around the boundary of any rectangle R the closure of which is in D.

Since D is simply connected, there exists on D a function $f(x,u)$, unique up to an additive constant, satisfying

$$f_x(x,u) = y(x;u), \qquad f_u(x,u) = -v(x;u). \tag{8}$$

Hence (2) is identical with (5), and (3) with (6). Clearly, f has the properties (α)-(δ).

In order to complete the proof of (I), the converse of the statement just proved must be verified. Let $f(x,u)$ have the properties (α)-(γ) on some domain D; so that T_2 and T_3 are continuous, one-to-one mappings of D onto (x,y)- and (u,v)-domains, respectively. Let (1) be the mapping $(x,y) \to (u,v)$ given by $T_1 = T_3 T_2^{-1}$. It must be shown that T_1 has the properties (i)-(iv). That T_1 has the property (i) is obvious; that T_1 has the property (ii) is equivalent to the fact that f has the property (β) (cf. (2) and (5)); that T_1 is orientation-preserving follows from the fact that (β) and (γ) imply that T_3 and T_2 are.

There only remains the verification of the statement that T_1 is area-preserving. By reversing the reasoning based on (7), it follows that T_1 preserves the area of any domain of the type of R_5. Since R_5 is the image under T_5 ($= T_2$) of a rectangle in the (x,u)-domain D, it is clear that the collection of sets R_5 forms a "basis" for the additive set-function which is the area. Hence T_1 is area-preserving. This proves (I).

2. Two lemmas. Condition (γ) in (I) does not imply (β). For the function $f(x,u) = x^3u$ has the partial derivative $f_u = x^3$, which is an increasing function of x, and the partial derivative $f_x = 3x^2u$, which is not an increasing function of u if $x = 0$. On the other hand, a weakened form of (γ) does imply a weakened form of (β).

LEMMA 1. *Let $f(x,u)$ be a function possessing continuous partial derivatives and let $f_u(x,u)$ be a non-decreasing function of x for fixed u. Then $f_x(x,u)$ is a non-decreasing function of u for fixed x.*

Proof. Let the segment $x = x_1$, $u_1 \leqq u \leqq u_2$ be in the domain D of definition of $f(x,u)$. It will be shown that $f_x(x_1,u_2) \geqq f_x(x_1,u_1)$. Let

x_2 $(> x_1)$ be chosen so that the rectangle $x_1 \leqq x \leqq x_2$, $u_1 \leqq u \leqq u_2$ is in D. Put

$$Q = f(x_1, u_1) - f(x_2, u_1) + f(x_2, u_2) - f(x_1, u_2).$$

The mean value theorem of differential calculus, applied to the function $f(x_2, u) - f(x_1, u)$ of u, shows that there exists a number u^* satisfying

$$Q = (f_u(x_2, u^*) - f_u(x_1, u^*))\,(u_2 - u_1) \text{ and } u_1 < u^* < u_2.$$

The assumption that f_u is a non-decreasing function of x implies that $Q \geqq 0$. On the other hand

$$0 \leqq Q/(x_2 - x_1) \to f_x(x_1, u_2) - f_x(x_1, u_1), \text{ as } x_2 \to x_1.$$

This proves Lemma 1.

It was noted in the proof of (I) that if (1) is a continuous, one-to-one mapping and $u(x, y)$ is an increasing function of y for fixed x, then, in the inverse mapping (4), the function $x(u, v)$ is strictly monotone in v for fixed u; cf. (ii′). A corresponding fact holds if $u(x, y)$ is only monotone (not strictly monotone).

LEMMA 2. *Let* (1) *be a continuous, one-to-one mapping and let* (4) *be its inverse. If* $u = u(x, y)$ *in* (1) *is a monotone function of* y *for fixed* x, *then* $x = x(u, v)$ *in* (4) *is a monotone function of* v *for fixed* u.

Proof. Since it is assumed that (1) is a continuous, one-to-one mapping of an (x, y)-domain onto a (u, v)-domain, the set of points $u(x, y) = \text{const.}$ is either empty or is an open Jordan arc. In the latter case, let $x = x(t)$, $y = y(t)$, where $0 < t < 1$, be a parametric representation of this arc. If two points (x_1, y_1), (x_1, y_2), where $y_2 > y_1$, are on this arc, then the entire segment $x = x_1$, $y_1 \leqq y \leqq y_2$ is on the arc, since u is a monotone function of y for fixed x. This shows that either $dx(t) \geqq 0$ or $dx(t) \leqq 0$ for $0 < t < 1$.

Since (1) is a topological mapping, $v = v(x, y)$ cannot have the same value at two distinct points of the arc $u = \text{const.}$ Hence either $dv(x(t), y(t)) > 0$ or $dv(x(t), y(t)) < 0$ for $0 < t < 1$. Thus, v can be used as a parameter on the arc $u = \text{const.}$ and x is a monotone function of v. This proves Lemma 2.

3. Semi-Legendre transformations. There will now be proved a theorem dealing with a semi-Legendre transformation of the type occurring in (2) and (3).

(II) *Let $z = z(x, y)$ possess continuous partial derivatives on some (x, y)-domain and let $z_x(x, y)$ be an increasing function of x for fixed y; so that*

$$T_9\colon\ p = z_x(x, y),\ y = y \tag{9}$$

is a one-to-one mapping $(x, y) \to (p, y)$. Then

$$g = xp - z, \tag{10}$$

as a function of (p, y), possesses the continuous partial derivatives

$$g_p = x, \qquad g_y = -z_y. \tag{11}$$

Proof. By virtue of T_9, any segment $p_1 \leqq p \leqq p_2$, $y = y_1$ in the (p, y)-plane corresponds in a one-to-one manner to a segment $x_1 \leqq x \leqq x_2$, $y = y_1$ in the (x, y)-plane. For any function f of (p, y) or (x, y), let $\Delta f = f(p_2, y_1) - f(p_1, y_1)$. Then $\Delta z = p^*\Delta x$, where $p^* = z_x(x^*, y_1)$ and $x_1 < x^* < x_2$, $p_1 < p^* < p_2$. Hence

$$\Delta g = p_1\Delta x + x_2\Delta p - \Delta z = x_2\Delta p + (p_1 - p^*)\Delta x < x_2\Delta p,$$

$$\Delta g = p_2\Delta x + x_1\Delta p - \Delta z = x_1\Delta p + (p_2 - p^*)\Delta x > x_1\Delta p.$$

Since $p_2 \to p_1$ implies $x_2 \to x_1$, it follows that $\Delta g/\Delta p \to x_1$ as $p_2 \to p_1$. Thus g_p exists and is x.

Let $p = p_1$ be fixed in such a way that $p(x, y) = p_1$ is a non-empty set of points (x, y). This set is a Jordan arc with a parametric representation $x = x(y)$. Let (x_1, y_1) and (x_2, y_2), where $y_2 > y_1$, be two points of this arc. For any function f, let $\Delta f = f(p_1, y_2) - f(p_1, y_1)$. Then Δz can be written in the following two ways: $\Delta z = p^*\Delta x + q^*\Delta y$ and $\Delta z = p^0\Delta x + q^0\Delta y$, where $p^* = z_x(x^*, y_1)$, $q^* = z_y(x_2, y^*)$ and $p^0 = z_x(x^0, y_2)$, $q^0 = z_y(x_1, y^0)$, and x^*, x^0 lie between x_1 and x_2, finally $y_1 < y^*, y^0 < y_2$. Since $p_1 = z_x(x_1, y_1) = z_x(x_2, y_2)$ and since $z_x(x, y)$ is an increasing function of x, it follows that $p_1 \leqq p^*$ and $p_1 \geqq p^0$ if $x_2 \geqq x_1$ and these inequalities are reversed if $x_2 \leqq x_1$. Hence

$$\Delta g = p_1\Delta x - \Delta z = -q^*\Delta y + (p_1 - p^*)\Delta x,$$

$$\Delta g = p_1\Delta x - \Delta z = -q^0\Delta y + (p_1 - p^0)\Delta x$$

show that $\Delta g/\Delta y$ is between $-q^*$ and $-q^0$. Consequently, g_y exists and is $-q = -z_y(x, y)$. This proves (II).

Remark. A part of assertion (II) remains true if $z_x(x, y)$ is a non-

decreasing (instead of an increasing) function of x for fixed y. In this case, (10) remains a function of (p, y). For if B denotes the set of points (p, y) which are images of points (x, y) under T_9, then the pre-image of a point (p^*, y^*) of B is a segment (open, closed, half-closed or a point) on the line $y = y^*$ in the (x, y)-plane. It is clear that g, as a function of (x, y), is constant on this segment, since $z(x, y^*)$ is linear in x and $z_x(x, y^*) = p^*$ is constant on that segment. Also, the first part of the proof of (II) shows that

$$g(p_2, y_1) - g(p_1, y_1) = (p_2 - p_1)x^* \tag{12}$$

holds for some point (x^*, y_1) on any segment of the line $y = y_1$ (in the (x, y)-plane) which is mapped onto the segment $p_1 \leqq p \leqq p_2$, $y = y_1$ under T_9

The name *semi-Legendre* transformation for (9) is justified not only in the obvious manner but also by the following corollary of (II):

(III) *Let $z = z(x, y)$ possess continuous partial derivatives. Let the mapping T_9: $(x, y) \to (p, y)$, given by* (9), *and the Legendre mapping T_{13}: $(x, y) \to (p, q)$, given by*

$$T_{13}\colon\ p = z_x(x, y),\ q = z_y(x, y), \tag{13}$$

be one-to-one. Then the Legendre transformation T_{13} is the product $T_{13} = T_{14}T_9$ of the (one-to-one) semi-Legendre transformations given by (9) *and*

$$T_{14}\colon\ p = p,\ q = -g_y(p, y), \tag{14}$$

where g is defined by (10).

In order to see that (14) defines a one-to-one transformation, it is sufficient to remark that $g_y(p, y) = -q(= -z_y)$ is a strictly monotone funnction of y for fixed p. For, in view of the assumptions that T_{13} is one-to-one and that $p(x, y)$ is increasing with x, it follows that q is a strictly monotone function of y for fixed p; cf. the remarks preceding Lemma 2.

4. Legendre transformations. The following theorem will be proved:

(IV) *Let $z = z(x, y)$ possess continuous partial derivatives on some (x, y)-domain and let the Legendre transformation T_{13}: $(x, y) \to (p, q)$ be one-to-one. In addition, suppose that either*

$$z_x(x, y) \text{ is a montone function of } x \text{ for fixed } y \tag{15}$$

or

(16) $z_x(x, y)$ is a monotone function of y for fixed x.

Then, as a function of (p, q),

$$h = xp + yq - z \tag{17}$$

has continuous partial derivatives; the transformation inverse to T_{13} *is the Legendre transformation*

$$T_{13}\colon\ x = h_p(p, q),\ y = h_q(p, q)\,; \tag{18}$$

finally, in the respective cases (15) *or* (16),

(19) $h_q(p, q)$ is a monotone function of q for fixed p

or

(20) $h_q(p, q)$ is a monotone function of p for fixed q.

Thus the relation $z \to h$ *is involutory.*

It is understood that the sense of the monotony in (15), (16), (19) and (20) is independent of y, x, p and q, respectively. It is also clear that if it is assumed that the monotony in (15) is strict, then the monotony in (19) is; if the monotony in (16) is strict, then (20) can be improved to "$h_p(p, q)$ is a strictly monotone function of q for fixed p."

A theorem corresponding to (IV) in which it is assumed that z is of class C^2 and that the Jacobian $\partial(p, q)/\partial(x, y) = z_{xx}z_{yy} - z_{xy}^2$ is not 0 is classical. Also, if it is assumed that z is convex (instead of assuming merely that z_x is monotone with respect to one of its arguments), then (IV) is known ([2]; for a similar theorem, see [3], pp. 365-366).

The particular case of (IV) in which it is assumed that $z_x(x, y)$ is a strictly monotone function of x for fixed y is contained in (III).

It will remain undecided whether or not (IV) remains true if the assumption on the monotony of z_x and the assertion on the monotony of h_q are omitted.

Proof of (IV). The theorem (IV) will first be proved under the assumption (16); say, $p = z_x(x, y)$ is a non-decreasing function of y for fixed x. By Lemma 1, the function $q = z_y(x, y)$ is non-decreasing in x for fixed y.

If the number $p = p_1$ is so chosen that the set of (x, y)-points $p(x, y) = p_1$ is not empty, then it is a Jordan arc. Let J be a closed subarc of this arc;

so that J has a parametric representation $x = x(t)$, $y = y(t)$, where $t_1 \leqq t \leqq t_2$ and $dx(t) \geqq 0$. It will be shown that

$$\int_J p\,dx + q\,dy = \Delta z, \tag{21}$$

where $\Delta z = z(x_2, y_2) - z(x_1, y_1)$ and $x_k = x(t_k)$, $y_k = y(t_k)$ for $k = 1, 2$. The arc J is not stated to be rectifiable, but the line (or Stieltjes) integral (21) exists, since p is constant and q is a monotone function of t along J.

Consider a subarc of J. For convenience, it will be supposed that the end points of this subarc are also denoted by (x_1, y_1), (x_2, y_2), where $x_2 \geqq x_1$. Then

$$\Delta z = p^*\Delta x + q^*\Delta y = p^0\Delta x + q^0\Delta y, \tag{22}$$

where $p^* = z_x(x^*, y_1)$, $q^* = z_y(x_2, y^*)$ and $p^0 = z_x(x^*, y_1)$, $q^0 = z_y(x_1, y^0)$, with $x_1 \leqq x^*$, $x^0 \leqq x_2$ and with y^*, y^0 between y_1 and y_2.

Since T_{13} is a one-to-one mapping, $q(t) = z_y(x(t), y(t))$ is a (strictly) monotone function of t. For the sake of definiteness, suppose that it is an increasing function. On the subarc under consideration, there is at least one value t^* of t satisfying $y(t^*) = y^*$ (and, of course, $x(t^*) \leqq x_2$). Since $q(x, y)$ is a non-decreasing function of x, it follows that $q^* \geqq q(t^*)$. Hence $q^* \geqq q(t^*) \geqq q(t_1)$. Similarly, $q^0 \leqq q(t_2)$.

Thus (22) shows that, if $\Delta y \geqq 0$,

$$\Delta z \geqq p(t_1)\Delta x + q(t_1)\Delta y + (p^* - p(t_1))\Delta x,$$

$$\Delta z \leqq p(t_2)\Delta x + q(t_2)\Delta y + (p^0 - p(t_2))\Delta x,$$

and similar inequalities hold if $\Delta y \leqq 0$.

Consider a mesh on the arc J and the corresponding inequalities for each subarc of the mesh. If the corresponding inequalities are added, there result the inequalities

$$\Delta z \geqq \int_J p\,dx + q\,dy + o(1) + o(1)\Sigma\,|\,\Delta x\,|,$$

$$\Delta z \leqq \int_J p\,dx + q\,dy + o(1) + o(1)\Sigma\,|\,\Delta x\,|,$$

where the $o(1)$-terms refer to the degree of fineness of the mesh tending to 0. The first $o(1)$ in each of the inequalities is the difference between the integral and a suitably chosen approximating sum $\Sigma(p_k\Delta^k x + q_k\Delta^k y)$. The second

$o(1)$ in each of the inequalities depends on the degree of continuity of $p = z_x(x, y)$. Since $dx(t) \geqq 0$ implies that $\Sigma \mid \Delta x \mid = x_2 - x_1$, the relation (21) follows from the last two inequalities.

An integration by parts, applied to (21), shows that

$$\Delta z = \Delta(xp + yq) - \int_J ydq \quad (dp = 0)\,; \text{ hence } \Delta h = \int_J ydq \quad (dp = 0),$$

by the definition (17) of h. Consequently, $\partial h/\partial q$ exists and is y.

Since the conditions on $z_x(x, y)$ and $z_y(x, y)$ are symmetric by virtue of Lemma 1, it follows that (21) holds if J is a closed subarc on which $q(x, y) = \text{const}$. Hence $\partial h/\partial p$ exists and is x.

Finally, if p, q are read in place of u, v in Lemma 2, it follows from (16) that $x = h_p(p, q)$ is a monotone function of q. Hence (20) is implied by Lemma 1. This proves the case (16) of (IV).

Assume that (15) holds; for instance, that $p = z_x(x, y)$ is a non-decreasing function of x for fixed y. If J is a closed arc on which $p(x, y) = p_1$, then J has a parametrization $x = x(t)$, $y = y(t)$, where $t_1 \leqq t \leqq t_2$ and $dy(t) \geqq 0$. By an argument similar to that used above, it will be shown that (21) holds.

First, the equations (22) hold, where x^*, x^0 are between x_1 and x_2 and where $y_1 \leqq y^*$, $y^0 \leqq y_2$. Since $p_1 = p(x_1, y_1) = p(x_2, y_2)$ and since $p(x, y)$ is a non-decreasing function of x for fixed y, it follows that $p^* \geqq p_1 \geqq p^0$ or $p^* \leqq p_1 \leqq p^0$ according as $x_2 \geqq x_1$ or $x_1 \leqq x_2$. Thus, if $x_2 \geqq x_1$,

$$\Delta z \geqq p(t_1)\Delta x + q(t_1)\Delta y + (q^* - q(t_1))\Delta y,$$

$$\Delta z \leqq p(t_2)\Delta x + q(t_2)\Delta y + (q^0 - q(t_2))\Delta y,$$

and similar inequalities hold if $x_2 \leqq x_1$. Since $\Delta y \geqq 0$, it can be concluded that (21) holds. Consequently, $\partial h/\partial q$ exists and is y.

It will now be shown that $\partial h/\partial p$ exists and is x. An integration by parts applied to the first term in (21) gives $\Delta z = \Delta(px) + \int_J qdy \quad (dp = 0)$. Since $dy \geqq 0$, it follows that

$$\Delta(px - z) = -q^*\Delta y \qquad (dp = 0), \tag{23}$$

where q^* is the value of $q = z_y$ at some point of J.

In view of the Remark following the proof of (II), it follows by (12) that the function $g = px - z$ of (p, y) satisfies

$$\Delta(px - z) = x^*\Delta p, \qquad (dy = 0), \tag{24}$$

where x^* is the abscissa of some point of a line segment $y = \text{const.}$ on which p varies from p_1 to p_2 (and $\Delta p = p_2 - p_1$).

Consider two points (x_1, y_1), (x_2, y_2) of an arc $q(x, y) = q_1$ and let $p_1 = p(x_1, y_1) < p_2 = p(x_2, y_2)$. If the two points are sufficiently close together, they can be joined by a path consisting of an arc on which $p = p_1$ and of a line segment on which $y = y_2$, and also by a path consisting of a line segment on which $y = y_1$ and of an arc on which $p = p_2$. Correspondingly, (23) and (24) give

$$\Delta(xp - z) = -q^*\Delta y + x^*\Delta p = x^0\Delta p - q^0\Delta y.$$

Since y is a monotone function of q for fixed p (that is, $dy \gtrless 0$ and $\Delta q \neq 0$ on the arcs $p = p_1$ and $p = p_2$), it follows that q_1 is between q^* and q^0. Hence $\Delta h = q_1\Delta y + \Delta(xp - z)$ is between $x^*\Delta p$ and $x^0\Delta p$. Consequently, $\partial h/\partial p$ exists and is x.

Since (19) follows from (15) and Lemma 2, this completes the proof of (IV).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] C. F. Gauss, *Werke*, vol. 8, p. 373.

[2] P. Hartman and A. Wintner, "On pieces of convex surfaces," *American Journal of Mathematics*, vol. 75 (1953), pp. 477-487.

[3] H. Lewy, "A priori limitations for solutions of Monge-Ampère equations. II," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 365-374.

[4] G. Scheffers, *Einführung in die Theorie der Flächen*, vol. 1 (ed. 1901), pp. 119-124.

ON THE NON-PERIODICITY OF THE ZEROS OF THE RIEMANN ZETA-FUNCTION.*

By C. R. PUTNAM.

In the half-plane $\sigma > 0$ of the complex $(z = \sigma + it)$ plane, the Riemann Zeta-Function, $\zeta(z)$, can be represented as

$$\zeta(z) = z \int_0^\infty ([x] - x) x^{-z-1} dx; \tag{1}$$

cf., e. g., [1], p. 14. The following theorem will be proved:

(*) *The sequence of the consecutive positive zeros of* $\zeta(\frac{1}{2} + it)$ *does not contain any periodic subsequence.*

Here, by a periodic sequence is meant any arithmetical progression of the form

$$d, 2d, 3d, \cdots, \qquad (d > 0). \tag{2}$$

(Needless to say, the assertion of (*) remains valid if the word "positive" is replaced by "negative.") The theorem states that every sequence (2) contains an infinity of elements which are not the imaginary parts of zeros of $\zeta(z)$ on the critical line $\sigma = \frac{1}{2}$. It should be noted that $\zeta(\frac{1}{2} + it)$ possesses an infinity of positive zeros $t_1 < t_2 < \cdots$ (Hardy); cf. [1], p. 214. Moreover, one has even $\liminf | t_{k+1} - t_k | = 0$, as $k \to \infty$. This follows from Selberg's result that $N(T) > \text{const. } T \log T$, where "const." denotes a positive constant and $N(T)$ denotes the number of zeros of $\zeta(\frac{1}{2} + it)$ on the interval $0 \leqq t \leqq T$; see [1], p. 226.

Actually, the theorem (*) holds if the role of the line $\sigma = \frac{1}{2}$ is played by any line $\sigma = \text{const.}$, whether or not the Riemann hypothesis is assumed. However, the assertion of (*) in this case is an easy consequence of the relation $N_S(T) = o(T)$ (Bohr-Landau), in which $S > \frac{1}{2}$ and $N_S(T)$ denotes the number of zeros $\sigma + it$ of $\zeta(z)$ in the rectangle $S \leqq \sigma \leqq 1$, $0 \leqq t \leqq T$.

Proof of (*). Another expression for $\zeta(z)$, easily deduced from (1), is given by

$$\zeta(z) = z(z-1)^{-1} - z \int_0^\infty g(x) e^{-itx} dx, \qquad 0 < \sigma < 1, \tag{3}$$

* Received July 11, 1953.

where, for $n = 1, 2, \cdots$,

$$g(x) = (e^x - [e^x])e^{-\sigma x} \text{ for } \log n \leqq x < \log(n+1). \tag{4}$$

If $z = \sigma + it$ is a zero of $\zeta(z)$, it follows from (3) that

$$(z-1)^{-1} = \int_0^\infty g(x)e^{-itx}dx,$$

and hence,

$$\int_0^\infty g(x)\cos(tx)dx = (\sigma-1)/((\sigma-1)^2+t^2). \tag{5}$$

In order to prove (*), suppose, if possible, that the assertion of the theorem is false; then there exists some number $d > 0$ such that, for $n = 1, 2, \cdots$, the numbers $z = \frac{1}{2} + idn$ are zeros of $\zeta(z)$. Henceforth, it will be supposed that $\sigma = \frac{1}{2}$. Next, extend the domain of definition of $g(x)$ to $-\infty < x < \infty$, by putting $g(-x) = g(x)$, $0 \leqq x < \infty$, and define $G(x)$ on $-\infty < x < \infty$ by

$$G(x) = \sum_{k=-\infty}^{\infty} g(x + 2\pi k/d). \tag{6}$$

In view of (4), one has $|g(x)| \leqq e^{-\frac{1}{2}|x|}$, $-\infty < x < \infty$, and it is clear that the series (6) is uniformly convergent on any finite interval. Moreover, $G(x)$ is periodic with period $2\pi/d$, so that $G(x + 2\pi/d) = G(x)$. In view of the relation

$$\int_0^{2\pi/d} G(x)e^{idnx}dx = \int_{-\infty}^{\infty} g(x)e^{idnx}dx,$$

(see [2], p. 37), it follows from (5), where now $\sigma + it = \frac{1}{2} + idn$, that, for $n = 1, 2, \cdots$,

$$\int_0^{2\pi/d} G(x)\cos(dnx)dx = -(\tfrac{1}{4} + d^2n^2)^{-1}$$

and

$$\int_0^{2\pi/d} G(x)\sin(dnx)dx = 0.$$

Thus, the Fourier coefficients of $G(x)$ are $O(n^{-2})$, hence $o(n^{-1})$. Therefore, the function $G(x)$ cannot possess a simple discontinuity; [2], p. 28.

Let $h > 0$. Since $g(x+0)$ $(= \lim g(x+h)$ as $h \to +0)$ exists for all x, it follows from the uniform convergence of the series (6) that $G(x+0)$ exists; similarly $G(x-0)$ exists. Next, it will be shown that $G(x+0) - G(x-0) < 0$ for at least one value of x, in contradiction with the assertion at the end of the last paragraph.

In order to show this, let $x = \log m$, where $m \geqq 2$ is an integer to be determined later. Then

$$G(\log m + h) - G(\log m - h) = g(\log m + h) - g(\log m - h)$$
$$+ \sum_{k \neq 0} [g(\log m + 2\pi k/d + h) - g(\log m + 2\pi k/d - h)]. \tag{7}$$

Since, by (4), $g(\log m + 0) - g(\log m - 0) < 0$, and since the jumps $g(x + 0) - g(x - 0)$, which occur only at the values $x = \pm \log n$, $n = 2, 3, \cdots$, are positive or negative according as $x = -\log n$ or $x = \log n$, it is clear that it is sufficient to show that m can be chosen so that

$$\log m + 2\pi k/d \neq -\log n \qquad (k = \pm 1, \pm 2, \cdots; n = 2, 3, \cdots). \tag{8}$$

(If (8) holds, then none of the terms of the summation of (7) can make a positive contribution as $h \to +0$.) Thus, it is enough to show that, for *some* integer $m \geqq 2$.

$$2\pi k/d \neq \log(mn) \text{ holds for } k = 1, 2, \cdots; n = 2, 3, \cdots. \tag{9}$$

Suppose though, if possible, that (9) fails to hold for every integer $m \geqq 2$. Then, if $m = m_j \geqq 2$ $(j = 1, 2)$ denote arbitrary integers, there exist integers $k = k_j \geqq 1$, $n = n_j \geqq 2$ such that $2\pi k_j/d = \log(m_j n_j)$ and hence $(m_2 n_2)^{k_1} = (m_1 n_1)^{k_2}$. However this last equality must fail to hold, a contradiction, provided only that m_1 and m_2 are chosen so that m_2 is relatively prime both to m_1 and to n_1 (note that n_1 depends only on m_1).

Hence (9) must hold for some integer $m = m' \geqq 2$ and therefore $G(x + 0) - G(x - 0) < 0$ for $x = \log m'$. As pointed out above, this leads to a contradiction and the proof of (*) is now complete.

PURDUE UNIVERSITY.

REFERENCES.

[1] E. C. Titchmarsh, *The Theory of the Riemann Zeta-Function*, Oxford (1951).

[2] A. Zygmund, *Trigonometrical Series*, Warsaw (1935).

A GENERAL THEORY OF RADICALS.*

II. Radicals in rings and bicategories.

By S. A. Amitsur.

In the introduction of part I ([1])[1] we have pointed out the relation between some of the well known radicals of rings and properties of rings which are invariant under homomorphisms (compare [3], Theorem 2). The object of the present paper is to give an axiomatic foundation of these radicals. If π is a property of rings, we generally mean by a *π-radical* of a ring R a unique maximal ideal N of R possessing the property π and such that the quotient R/N does not contain non-zero π-ideals. The main properties which a radical is generally required to satisfy is: the existence in every ring; the radical of an ideal A should be the intersection of A and the radical of the whole ring; the radical should contain also one-sided ideals; the radical of a matrix ring over a ring R should be the matrix ring over the radical of R. Necessary and sufficient conditions for these results to be satisfied by a property π are given, among which the most important is the fact that π should be invariant under homomorhpism. If only this condition is imposed on π (and another of far less importance) one can obtain generalizations of Baer's Lower Radical ([5]). Some methods of constructing properties which yield radicals from non-yielding-radicals properties are given in section 7 and 8. This method will be applied to prove some new results about some known radicals, in particular Baer's Radical.

The most important tools in studying this homomorphic-type radicals are undoubtedly the main two Isomorphism Theorems. This is the reason why the whole theory (except some particular ring-structure problem, e. g. radicals of matrix rings) can be developed in quite a general class of bicategories, namely: the lattice ordered bicategories, in the sense of MacLane ([7]), which contain a zero element and which satisfy some additional axioms. In particular, the theory will hold for a big class of abstract algebras containing rings (associative or non-associative), semi-groups, groups, loops, etc. The

* Received September 10, 1952; substantially revised March 7, 1953.

[1] The results of part I ([1]) will be quoted by the number I followed by the notation of the result quoted, e. g. definition I. 1 mean definition 1 of the paper [1].

greatest part of the present paper deals only with rings (not necessarily associative) since the main applications will be for the theory of rings. Nevertheless, in section 8 we give a list of conditions to hold in bicategories so that the theory developed in the preceding section will hold for them.

The last section is devoted to the study of the relation between the theory of radicals as developed by B. Brown and N. H. McCoy in [3] and [4] and the present theory. The type of radical dealt with by these authors constitutes an important class of the radicals whose theory is developed here. The incorporation of their theory and the present theory yields a slight extension of the results of [3] but nevertheless it makes use of Zorn's lemma which was avoided in [3].

1. *HI*- and *RI*-properties. In the present section we restate some of the main results of part I ([1])[1] in the language of ideals and homomorphisms of rings. This is obtained by applying the results of that paper to the lattice of all ideals of rings. These restated results are needed for further applications and for a simplification of the theory developed in the preceding paper when applied to rings.

Let π be a property of ideals in rings. By a *π-ideal* we mean an ideal which possesses the property π.

The sequence of ideals $U_\lambda = U_\lambda(R)$ is defined as follows:

1) $U_0 = 0$.

2) $U_\lambda = \bigcup_{\nu<\lambda} U_\nu$ for limit ordinals λ.

3) For non limit ordinals λ, U_λ is the union of all the ideals A of R such that $A/U_{\lambda-1}$ is a π-ideal in $R/U_{\lambda-1}$.

The ideal U_λ is known as the λ-th (π-) radical of R The *upper π-radical* $U(R,\pi) = U(R)$ of R is defined to be the limit ideal of this sequence; namely the minimal ideal U_λ such that $U_\lambda = U_{\lambda+1}$. The following notions are fundamental:

π-semi simple: a ring R which does not contain non-zero π-ideals.

$\overline{\pi}$-ideal: an ideal P in a ring R for which the quotient R/P is π-semi simple.[2]

π-radical: an ideal N in a ring R which is both a π- and a $\overline{\pi}$-ideal in R. That is, N is a π-ideal and the quotient R/N is π-semi simple.

---

[2] These ideals correspond to radical ideals in the sense of Baer [5].

A property π of ideals will be called an *HI-property* if π satisfies the following two conditions:

(1A) *The zero is a π-ideal.*

(1B) *Every homomorphism maps π-ideals onto π-ideals.*

For a property π of ideals, we define a relation π in the lattice of all ideals of a ring R as follows: let A and B be two ideals in R. Using the notation of part I ([1]), we write $A \pi B$ if $A \supseteq B$ and A/B is a π-ideal in R/B. Since the lattice of all ideals of a ring is known to be complete, the results of part I ([1]) can be applied to this lattice in view of the following fundamental lemma:

LEMMA 1.1. *If π is an HI-property then the relation π defined in the lattice of the ideals is an H-relation in the sense of definition* I. 1.[1]

Indeed, conditions I. (A) and I. (B) of [1] are evidently satisfied. To prove I. (C), we consider three ideals A, B, C in R such that $A \pi B$ and $C \supseteq B$. The homomorphic mapping of R/B onto R/C maps A/B onto $A \cup C/C$; hence by (1B) it follows that $A \cup C/C$ is a π-ideal in R/C. Thus $(A \cup C)\pi C$, q. e. d.

The application of Theorems I. 1. 2, I. 1. 1, I. 1. 3, and Corollary I. 1. 4 yields

THEOREM 1.1. *If π is an HI-property then the intersection of any set of $\overline{\pi}$-ideals of a ring R is a $\overline{\pi}$-ideal; the intersection of all $\overline{\pi}$-ideals of R, which is the minimal $\overline{\pi}$-ideal, is exactly the upper π-radical $U(R)$ of R. Furthermore, if a π-radical N of R exists, then $N = U(R)$ and N is uniquely determined as the maximal π-ideal of R.*

In the latter case N is called *the π-radical* of R and will be denoted by $\pi(R)$.

Properties for which the π-radical always exists must satisfy:

(1C) *For every ideal A in a ring R which is not a π-ideal, there exists an ideal $B \subseteq A$ such that there are no non zero π-ideals in R/B contained in A/B.*

Properties which satisfy the three conditions (1A), (1B) and (1C) will be called *RI-properties*, and for these properties we have by Theorem I. 2. 1:

THEOREM 1.2. *Let π be an HI-property, then a necessary and sufficient condition that the π-radical exists in every ring is that π be an RI-property.*

As a consequence of Theorem I. 2. 2 we obtain the following fundamental result concerning *RI*-properties:

THEOREM 1. 3. *The union of any set of π-ideals is a π-ideal.*

The conditions of Theorem I. 2. 3 can be simplified in the present application as follows

THEOREM 1. 4. *If π is an HI-property then the following two conditions are equivalent to* (1C)

($1C_1$) *If $A \supseteq B$ are ideals in R such that B is a π-ideal in R and A/B is a π-ideal in R/B, then A is also a π-ideal in R.*

($1C_2$) *If $A_1 \subseteq A_2 \subseteq \cdots$ is a non decreasing well ordered sequence of π-ideals then the union $\cup A_\nu$ is also a π-ideal.*[3]

This theorem is a consequence of Theorem I. 2. 3 since ($1C_1$) implies I. (D_1) and ($1C_2$) implies I. (D_2). Indeed, if $A \pi B$ and $B \pi C$ in the lattice of ideals of R then, by (1B), $(A/C)\pi(B/C)$ in the lattice of ideals of R/C since $(A/C)/(B/C) \cong A/B$. Applying now ($1C_1$) to the ring R/C we obtain $A \pi C$, which proves I. (D_1). Similarly, if $B \subseteq A_1 \subseteq A_2 \subseteq \cdots$ is a sequence of ideals in R such that $A_\nu \pi B$ then applying ($1C_2$) to the ring R/B we see that $\cup (A_\nu/B)$ is a π-ideal. Now, $\cup (A_\nu/B) = (\cup A_\nu)/B$. It follows, therefore, that $(\cup A_\nu)\pi B$ which proves I. (D_2).

The homomorphism invariance of π immediately yields the following results:

THEOREM 1. 5. 1) *Let ϑ be an isomorphism between R and R^ϑ, then ϑ induces a one to one correspondence between the π-ideals of R and R^ϑ.*

Hence,

2) *The isomorphic images of π-semi simple rings are π-semi simple.*

We are now able to prove:

THEOREM 1. 6. *A subdirect sum of π-semi simple rings is π-semi simple.*

Indeed, if R is a subdirect sum of π-semi simple rings $\{R_\nu\}$ then $0 = \cap A_\nu$ where A_ν are ideals in R such that $R_\nu \cong R/A_\nu$. From the pre-

[3] In the proof of the equivalence of these conditions the well-ordering theorem is used. One can use Zorn's lemma instead by replacing ($1C_2$) by the requirement that the union of every linear system of π-ideals is a π-ideal.

ceding theorem it follows that $A_\nu \bar{\pi} R$ [4] in the lattice of ideals of R. Hence, by Lemma I. 1. 3 it follows that $0 \bar{\pi} R$ which evidently means that R is π-semi simple.

Thus, Theorem I. 1. 5 yields:

THEOREM 1. 7. *If the kernel ϑ of the homomorphic mapping of R onto R^ϑ is contained in the upper π-radical $U(R)$, then $U(R^\vartheta) = U(R)^\vartheta$.*

In particular, if ϑ is an automorphism we obtain:

COROLLARY 1. 1. *$U(R)$ is invariant under the group of automorphisms of R.*

Example. For associative rings, properties like regularity, strong regularity, and π-regularity [5] are HI-properties. The property of quasi-regularity is an example of an RI-property.

2. Ring properties. The rest of this paper deals with a less general class of properties. We consider henceforth only properties of rings (not of ideals in rings!); thus nilpotency of associative rings is such a property.

By a *π-ring* R we shall mean a ring R which possesses the property π. A ring-property π is turned into a property of ideals in rings by defining a *π-ideal* to be an ideal which is also a *π-ring*.

In contrast with the properties of ideals considered in the preceding section, the ring properties π have the characteristic that if A is a π-ideal in some ring R then A remains a π-ideal in every ring containing A as an ideal. This fact yields some modification of the conditions of the preceding section.

For the convenience of the reader we give here a list of all conditions which will be used in the present paper:

(A) The zero is a π-ring.

(B) Homomorphic images of π-rings are π-rings.

(C) Every non π-ring is homomorphic to a non zero π-semi simple ring.[6]

---

[4] The notation $\bar{\pi}$ is used here in the sense of section I. 1. That is: $B \bar{\pi} A$ means that $A \supseteq B$ and the quotient A/B does not contain non zero π-ideals of R/B.

[5] For definition see e. g. I. Kaplansky, "Topological representation of algebras II," *Transactions of the American Mathematical Society*, vol. 68 (1950), p. 67.

[6] This amounts to the fact that every ring which is not a π-ring properly contains a $\bar{\pi}$-ideal.

(D) All ideals of a π-ring are π-ideals.

(E) All ideals of a π-semi simple ring are π-semi simple rings.

For simplification we introduce the following names: properties which satisfy (A), (B) will be called *H-properties* (homomorphism invariant properties). If (A), (B), (C) are satisfied, the property will be called an *R-property* (a radical-property). *H*-properties which satisfy (D), (E) will be called *SH-properties.* If all conditions are valid, the property will be called an *SR-property.*

The following two conditions will be seen to be equivalent to (C):

(C_1) (Transitivity) If A is an ideal in R such that R/A and A are π-rings then R is also a π-ring.

(C_2) If $A_1 \subseteq A_2 \subseteq \cdots$ is a well-ordered sequence of π-ideals then $\cup A_\nu$ is a π-ideal.

For problems relating to one sided ideal and subrings we shall need the following conditions (section 3):

(D_r) Every right ideal of a π-ring is a π-ring.

(F_r) π-semi simple rings do not contain non zero right π-ideals.

(D_s) Every subring of a π-ring is a π-ring.

(F_s) π-semi simple rings do not contain non zero π-subrings.

For determining the radical of a matrix ring R_n over a ring R, we shall need the following conditions (section 4):

(G_1) If R is a π-ring then so is R_n.

(G_2) If R is π-semi simple then so is R_n.

Two more conditions of a less general nature will appear in sections 6 and 8.

The classes of *H*-properties and *R*-properties defined above are subclasses of the *HI*-properties and *RI*-properties respectively as will be shown in the following lemma:

LEMMA 2.1. 1) *An H-property is an HI-property.*

2) *R-properties are also RI-properties.*

Before proving the lemma, we note that (C_1) is readily seen to be equivalent to condition ($1C_1$) of the preceding section, and (C_2) is exactly the condition ($1C_2$).

The proof of (1) is evident. To prove (2) we shall show that if π is an H-property, then a necessary and sufficient condition for the existence of the π-radical is that π satisfies (C); and thus the proof of (2) will follow by Theorem 1. 2. To this end we observe that if π is an H-property whose π-radical always exists then the quotient ring modulo this radical satisfies (C); and hence π is an R-property. Now suppose that π is an R-property. Let $A \supseteq B$ be two ideals in a ring R such that A/B is a π-ring in R/B and B is a π-ring in R. Consider A as a ring; if it were not a π-ring then, by (C), A would properly contain a π-ideal D. Since B is a π-ideal it follows, by Lemma I. 1. 2, that $D \supseteq B$. In view of the fact that $(A/B)/(D/B) \cong A/D$ it follows that D/B is also a π-ideal in A/B; but A/B is a π-ring, hence $D \supseteq A$. This is a contradiction, which proves that π satisfies $(1C_1)$. To prove that π satisfies $(1C_2)$, we consider a well ordered non decreasing sequence of π-ideals $A_1 \subseteq A_2 \subseteq \cdots$. If $A = \cup A_\nu$ is not a π-ideal, then considering A as a ring we would obtain, by (C), the existence of a π-ideal D properly contained in A. Since A_ν is also an ideal in A, it follows by Lemma I. 1. 2 that $D \supseteq A_\nu$. Hence $D \supseteq \cup A_\nu = A$ which is a contradiction. The rest of the proof follows now immediately by Theorem 1. 4.

An immediate consequence of this theorem is that H-properties π yield lower π-radicals and R-properties yield π-radicals, which have the properties of the radicals of the preceding section.

The ring-properties satisfy also the following theorem:

THEOREM 2. 1. *If π is an H-property, then $U(S) \supseteq U(R)$ for every subring S which contains $U(R)$.*

The proof will follow by showing inductively that $U(S) \supseteq U_\lambda(R)$, where $U_\lambda(R)$ is the λ-th π-radical of R. Indeed, since $0 \subseteq S$, $U(S) \supseteq U_0(R)$. Suppose $U(S) \supseteq U_\nu(R)$ for every $\nu < \lambda$. For a limit ordinal λ, it is evident that this implies that $U(S) \supseteq U_\lambda(R)$.

If λ is a non limit ordinal and $P/U_{\lambda-1}(R)$ is a π-ideal in $R/U_{\lambda-1}(R)$, then since $S \supseteq U(R) \supseteq U_\lambda(R) \supseteq P$, $P/U_{\lambda-1}(R)$ is also a π-ideal in $S/U_{\lambda-1}(R)$; hence $U(S) \supseteq P$. By the definition of $U_\lambda(R)$ it follows now that $U(S) \supseteq U_\lambda(R)$.

In particular, if $S = U(R)$ we have:

COROLLARY 2. 1. $U(U(R)) = U(R)$.

Consider now the set of all subrings of a ring R. This set is turned into a complete pseudo-lattice, in the sense of definition I. 4, by defining a

relation: $A \geqq B$, between subrings, if B is an ideal in the subring A.[7] The intervals $[B, A]$ of this complete pseudo-lattice can be indentified with the complete lattice of the ideals of the quotient A/B. Let π be an H-property; as before, this property determines a relation π in this pseudo-lattice. Condition I. (C_1) of section I. 3 is readily seen to be a consequence of the first isomorphism theorem and Theorem 1. 5. Thus π determines an H-relation in the complete pseudo-lattice. Stating Theorem I. 3. 1 for the present case we have

THEOREM 2. 2. *Let π be an R-property. A necessary and sufficient condition that $\pi(A) = A \cap \pi(R)$ for every ideal A is that π be an SR-property, i. e. π should satisfy the five conditions* (A), (B), (C), (D) *and* (E).

The first isomorphism theorem together with (B) imply that SR-properties satisfy the conditions of Theorem I. 4. 1. Hence,

COROLLARY 2. 2. *The mapping $A \to r(A)$, where $r(A)/A$ is the π-radical of R/A, is an idempotent meet-homomorphism of the lattice of ideals of R.*

Another result is that H-properties which satisfy (D), satisfy the conditions of Theorem I. 5. 2. Thus, each of these properties determines a *dual radical.*

We conclude this section with proving the necessity of the conditions (A)-(E) for the existence of the radical. That is:

THEOREM 2. 3. *Let π be an isomorphism invariant property of rings. Then:*

1) *If every algebra R contains a maximal π-ideal $\pi(R)$ such that $\pi(R)$ is the minimal ideal P in R with the property that R/P is π-semi simple, then π is an R-property.*

2) *If the preceding ideal $\pi(R)$ satisfies: $\pi(A) = A \cap \pi(R)$ for every ideal A in a ring R then π is an SR-property.*

Proof. Since $\pi(0) = 0$ for the zero ring, it follows that π satisfies (A). To prove (B), let R be a π-ring and let R' be a homomorphic image of R. Denote by N the origin of $\pi(R')$ in R. Note that the isomorphism invariance of π implies that isomorphic images of π-semi simple rings are also π-semi simple. Applying this fact to the case $R/N \cong R'/\pi(R')$, where the latter is π-semi simple by the condition of the theorem, we obtain that R/N is π-semi

[7] This is the pseudo-lattice given as an example in section I. 3.

simple. Hence, by the assumption of the theorem, $N \supseteq \pi(R)$. Since R is a π-ring and $\pi(R)$ is the maximal π-ideal of R, $\pi(R) = R = N$, which proves that $N' = \pi(R') = R'$. Thus R' is a π-ring and (B) is proved. By now π has been proved to be an H-property, hence by the proof of Lemma 2.1 it follows that π is an R-property. The rest of the theorem follows now immediately from Theorem 2.2.

3. Subrings and one-sided ideals. The present section deals with the relation between the upper π-radical $\pi(R)$ of a ring R and its one sided ideals. We shall deal only with right ideals.

THEOREM 3.1. *If π is an H-property of rings, then the upper π-radical $U(R)$ of every ring R contains every right π-ideal of R if and only if π satisfies*:

(F_r) *Every π-semi simple ring does not contain non zero right π-ideals.*

Put $U = U(R)$. Let J be any right π-ideal of R. $J/J \cap U$ is a homomorphic image of J, hence it is a π-ring. This immediately implies that $J \cup U/U$ is a π-ring, and, therefore, a right π-ideal in R/U. The π-semi simplicity of the latter and (F_r) shows that $U \supseteq J$.

The necessity part of this theorem is a simple consequence of the application of the result to the case $U(R) = 0$.

Another result for these properties is:

THEOREM 3.2. *If π is an R-property satisfying (F_r) then the union of any set of right π-ideals is also a right π-ideal.*

Let $\{J_\nu\}$ be a set of right π-ideals of R. Put $J = \cup J_\nu$; each J_ν is also a right π-ideal in J, hence $\pi(J) \supseteq J_\nu$. This implies $\pi(J) \supseteq \cup J_\nu = J$ which evidently means that J is a π-ring.

Thus,

COROLLARY 3.1. *If π is an R-property satisfying (F_r) then $\pi(R)$ is the unique maximal right π-ideal of R.*

The same methods yield similar results for left ideals under a similar Condition (F_l) imposed on left ideals. Furthermore, the same methods yield similar results for subrings. Namely:

THEOREM 3.3. 1) *If π is an H-property which satisfies*:

(F_s) *Every π-semi simple ring does not contain non zero π-subrings, then $U(R)$ contains every π-subring of R.*

2) *If π is an R-property of this type, then the union of every set of π-subrings is a π-subring and $\pi(R)$ is the maximal π-subring of R.*

In some cases the following is useful in proving the validity of (F_r). Denote by $\mathfrak{L}(R)$ the ring of all linear transformations of R: $a_l: x \to ax$, $a \in R$. Then:

LEMMA 3.1. *A property π satisfies* (F_r) *if it satisfies the requirement that for any right π-ideal J in R, the two sided ideal $\mathfrak{L}(R)J$ is a π-ideal. This requirement is also necessary if π is assumed to satisfy* (D).

Indeed, if the requirement of this theorem is satisfied and R is π-semi simple, $\mathfrak{L}(R)J = 0$. This means that J is two sided, and since it is a π-ideal, it follows that $J = 0$, q. e. d. The necessity part follows immediately by (D) and by the fact that $\pi(R) \supseteq J$ implies $\pi(R) \supseteq \mathfrak{L}(R)J$.

4. The radical of matrix rings. The present section deals with the relation of a radical of a ring R and the radical of a matrix ring over R. We use the notation R_n for the ring of all square matrices of order n on R, and R_f for the ring of finite-rowed infinite matrices (or finite columns). We prove:

THEOREM 4.1. *Let π be an H-property of rings. Then $U(R_n) = U(R)_n$ for every ring R if π satisfies the following conditions:*

(G_1) *If R is a π-ring then so is R_n.*

(G_2) *If R is a π-semi simple ring then so is also R_n.*

Proof. Consider the sequence $U_\lambda = U_\lambda(R)$ of the λ-th π-radicals. Put $U_{\lambda n} = (U_\lambda)_n$. Our first step in the proof is to show inductively that $U(R)_n \supseteq U_{\lambda n}$. For $\lambda = 0$, this is evident. If λ is a limit ordinal, $U(R)_n \supseteq U_{\lambda n}$ is a simple consequence of the induction and of the fact that $U_{\lambda n} = (\bigcup_{\nu<\lambda} U_\nu)_n = \bigcup_{\nu<\lambda} U_{\nu n}$. Let λ be a non limit ordinal and assume that $U(R) \supseteq U_{\lambda-1\, n}$. If P is an ideal in R such that $P/U_{\lambda-1}$ is a π-ideal, then by (G_1) it follows that $(P/U_{\lambda-1})_n$ is a π-ring. Since π is an H-property this implies that $P_n/U_{\lambda-1\, n}$ is a π-ideal in $R_n/U_{\lambda-1\, n}$; hence, by Theorem 1.7, it is readily seen that $U(R_n) \supseteq P_n$. This being valid for every such P yields that $U(R_n) \supseteq \cup P_n = (\cup P)_n = U_{\lambda n}$, q. e. d. It follows now by Theorem 1.1 that $U(R_n) \supseteq U(R)_n$. Since $R_n/U(R)_n \cong (R/U(R))_n$ and $R/U(R)$

is π-semi simple it follows by (G_2) and by Theorem 1. 1 that $U(R)_n \supseteq U(R)_n$. Thus $U(R_n) = U(R)_n$.

Remark 4. 1. If π is an R-property then (G_1) and (G_2) are also necessary conditions for the validity of the preceding theorem, since then (G_1) is the assertion of this theorem for $\pi(R) = R$ and (G_2) is the assertion for $\pi(R) = 0$.

As an immediate consequence of (G_1) and of the fact that $\pi(R)_n = \pi(R_n) = R_n$ for π-rings R, we obtain:

COROLLARY 4. 1. *If π is an R-property which satisfies (G_1) and (G_2) then R_n is a π-ring if and only if R is a π-ring.*

The converse of this corollary is generally not true. But suppose π is an R-property which satisfies (D) and the condition of the preceding corollary; then (G_2) must hold. If not, there exists a π-semi simple ring R such that R_n contains a non zero π-ideal A. Consider the ideal in R_n generated by the matrices $p(aq)$, $(pa)q$, $p, q \in R_n$ and $a \in A$. This ideal is contained in A, hence by (D) it is a π-ideal. One readily observes that this ideal equals P_n, where P is the ideal in R generated by the elements: $r(st)$, $(rs)t$, where $t, r \in R$ and s range over all the elements of R appearing in the matrices of A. Since R is π-semi simple it follows by the condition of the corollary that $P = 0$. In particular this yields:

THEOREM 4. 2. *If π is an R-property which satisfies* (D) *and the condition that a π-semi simple ring R cannot contain a non zero ideal P such that $R(PR) = (RP)R = 0$, then (G_1) and (G_2) are equivalent to the condition that R_n is a π-ring if and only if R is such.*

Note that in the whole section no use was made of the relation between the two operations of the rings. Hence the preceding results hold also for non associative and non distributive rings.

5. Dependence of conditions. The conditions listed in section 2 are generally independent, but not if one restricts oneself to particular classes of rings. The object of the present section is to show that in the case of associative rings condition (E) is superfluous.

LEMMA 5. 1. *Let π be an R-property of rings (not necessarily associative). If Q is an ideal in a π-semi simple ring R such that $Q^2 = 0$ then Q is also π-semi simple.*

Proof. Denote by $\mathfrak{T}(R)$ the ring of all transformations of the additive group of R generated by the identity and by the transformations $a_r\colon x \to xa$, $a_l\colon x \to ax$, $a \in R$. Suppose that Q contains a π-ideal P *in* Q. For every $\tau \in \mathfrak{T}(R)$ the totality of the elements $p\tau$, $p \in P$ constitutes an additive subgroup of Q. We note that since $Q^2 = 0$, the ideals of Q are its additive subgroups. Now the mapping $p \to p\tau$ defines a homomorphism of the additive group P onto $P\tau$ and, evidently, this is also a homomorphism of the ideal P onto the ideal $P\tau$; hence condition (B) implies that $P\tau$ is also a π-ideal in Q. By Theorem 1. 3 it follows that the union P^* of all the ideals $P\tau$, $\tau \in \mathfrak{T}(R)$, is a π-ideal in Q and, therefore, it is also a π-ring. The set P^* is readily seen to be an ideal in R. Thus the π-semi simplicity of R implies that $P^* = 0$ which proves that $P = 0$, q. e. d.

THEOREM 5. 1. *If π is an R-property of associative rings which satisfies condition* (D), *then π satisfies also* (E), *i. e. π is an SR-property.*

Proof. Let R be a π-semi simple ring and let P be an ideal in R. If Q is a π-ideal in P then $PQP \subseteq Q$, and it is a π-ideal in R by (D). By the π-semi simplicity of R it follows that $PQP = 0$. Let $P_r = \{x \mid x \in P, Px = 0\}$; then P_r is an ideal in R such that $P_r^2 = 0$. Since $QP \subseteq Q \cap P_r \subseteq Q$, it follows by (D) that $Q \cap P_r$ is a π-ideal in P_r.

Thus the preceding lemma implies that $P_r \cap Q = QP = 0$. Let $P_l = \{x \mid x \in P, xP = 0\}$ then P_l is an ideal in R such that $P^2{}_l = 0$. Since $Q \subseteq P_l$ the preceding argument yields that $Q = 0$ which proves (E).

Remark 5. 1. If one restricts oneself to associative rings possessing operators and considers only admissible ideals, then the same proof is valid also in this case, with the additional remark that in the proof of the first lemma one should note that if P is admissible each of the ideals $P\tau$ and also P^* is admissible. In the proof of the preceding theorem, one has to note that together with P and Q, also PQP, QP, P_r and P_l are admissible ideals.

Remark 5. 2. At the end of this paper we point out algebraic structures to which the theory of radicals can be extended, except for the parts which depend on ring operations. It is worth remarking at this stage that the theory can be developed without any changes for semi-groups (with or without operators) and in particular this section will hold for associative semi-groups.

6. Z-properties. Many of the properties of rings which yield radicals are properties which satisfy the following condition:

(Z) *Every zero ring, i. e. a ring R such that $R^2 = 0$, is a π-ring.*

A great deal of the theory can be simplified if (Z) is assumed. This is carried out in this section.

H-properties and R-properties which satisfy (Z) will be referred to as *HZ-properties* and *RZ-properties* respectively.

Condition (Z) itself can be chosen as a definition of a new ring property in the following way: a ring R is a ζ-ring if $R^2 = 0$. This property is evidently an HZ-property and yields, therefore, an upper ζ-radical. For asociative rings this radical is well known as Baer's Lower Radical defined in [5]. Additional results on this radical will be given in a subsequent paper.

THEOREM 6.1. *If π is an HZ-property, then the upper π-radical $U(R) \supseteq \zeta(R)$.*

This is an immediate consequence of Theorem I. 1. 6.

One readily verifies that the converse of the preceding theorem holds for R-properties π. That is: if $\pi(R) \supseteq \zeta(R)$ for every ring R then π is an RZ-property. The RZ-properties satisfy the following famous property of radicals.

THEOREM 6.2. *If π is an HZ-property of associative rings, then an element $r \in R$ belongs to $U(R)$ if and only if $RrR \subseteq U(R)$.*

Proof. Let $Q = \{x \mid x \in R, RrR \subseteq U(R)\}$. Then Q is an ideal in R such that $Q^3 \subseteq U(R)$. The natural homomorphism of R onto $R/U(R)$ maps Q on a nilpotent ideal of $R/U(R)$. By (Z) it is readily seen that π-semi simple rings do not contain nilpotent ideals, hence $Q \subseteq U(R)$.

The property (Z), when satisfied together with conditions (D_r) and (F_r) of section 4, yields the existence of the radical of a matrix ring:

THEOREM 6.3. *Let π be an RZ-property which satisfies (D_r)[8] and (F_r); then $\pi(R_n) = \pi(R)_n$.*

Proof. Let R be a π-ring and let $\{c_{ik}\}$ be a matrix unit base of R_n. Denote by $R^{(1)}$ the ring of all one rowed matrices: $\{r_1c_{11} + \cdots + r_nc_{1n}\}$. This ring contains the two sided ideal P of all the matrices $r_2c_{12} + \cdots + r_nc_{1n}$, and $P^2 = 0$. The quotient $R^{(1)}/P \cong Rc_{11}$ and the latter is isomorphic with R. Since R is a π-ring and P is a π-ideal, by (Z), it follows by (C_1) that $R^{(1)}$ is also a π-ring; hence $R^{(1)}$ is a right π-ideal in R_n. Similarly each of the right ideals $R^{(1)} = \{r_1c_{i1} + \cdots + r_nc_{in}\}$, $i = 1, \cdots, n$ is a right π-ideal in R_n.

[8] By Theorem 5. 1 it follows that π is also an *SRZ*-property.

Since R_n is the sum of all these ideals, it follows by Theorem 3. 2 that R_n is a π-ring, which proves (G_1) of section 4. Conversely, let R_n be a π-ring. By (D_r) it follows that the right ideal $R^{(1)}$ is a π-ring. As it has already been pointed out the latter is homomorphic with R, hence by (B), R is also a π-ring. The theorem follows now immediately by Theorem 4. 1 in view of the facts that RZ-properties satisfy Theorem 4. 2 and that (D_r) implies (D).

One readily observes that the theorems used above do not require the finiteness of n and we may similarly obtain:

THEOREM 6. 4. *If π is an RZ-property which satisfies (D_r) and (F_r) then $\pi(R)_f = \pi(R_f)$, where, e. g. R_f denotes the ring of all finite matrices over R or the ring of all finite rowed matrices over R.*

THEOREM 6. 5. *Let π be an RZ-property of associative rings which satisfies (D_r) and (F_r). Then the π-radical of a right ideal J is the set $J_r = \{x \mid x \in J, xJ \subseteq \pi(R)\}$.*

Proof. Evidently J_r is an ideal in J. Since $JJ_rJ \subseteq \pi(R)$ it follows by (D_r) that JJ_rJ is a right π-ideal; hence $JJ_rJ \subseteq \pi(J)$. It follows now, by Theorem 6. 2, that $J_r \subseteq \pi(J)$. Now $\pi(J)$ is a two sided ideal in J, hence by (D) we obtain that $\pi(J)J$ is a π-ring. This means that $\pi(J)J$ is a right π-ideal in R, which implies by Theorem 3. 1 that $\pi(J)J \subseteq \pi(R)$; hence $\pi(J) \subseteq J_r$, thus $\pi(J) = J_r$.

In particular, if R is π-semi simple we obtain:

COROLLARY 6. 1. *The π-radical of a right ideal in a π-semi simple ring is its left annihilator.*

A short cut in the proof of existence of the π-radical can be obtained for properties of associative rings which satisfy (D) with the aid of the following theorem:

THEOREM 6. 6. *If π is an HZ-property which satisfies (D) and the condition that every non π-ring is homomorphic to a non zero ring which does not contain non zero right π-ideals, then π is an RZ-property which satisfies (F_r).*

Evidently the condition of this theorem implies the validity of (C) which proves that π is an RZ-property. To prove (F_r), let R be a π-semi simple ring and let J be a right π-ideal in R. If $J \neq 0$, $J_0 = (J, RJ)$ is a non zero ideal in R which is not a π-ideal. Therefore, J_0 is homomorphic with a non zero ring J_0/Q which does not contain right π-ideals. Thus

$J \subseteq Q$. This yields $J^2_0 = (J^2, JRJ, RJ^2, (RJ)^2) \subseteq Q$, since $J_0 J \subseteq Q$ and $(RJ)^2 \subseteq (RJ)J \subseteq J_0 J \subseteq Q$. Hence J_0/Q is a π-ring. Contradiction!

7. Constructing new properties. Let π be any property of rings. The only restriction imposed is that the zero ring should be a π-ring, namely, that π satisfies (A). This property gives rise to an H-property by defining:

DEFINITION 7.1. *A ring R is to be called a π_h-ring if every homomorphic image of R (including R) is a π-ring.*

One readily observes that π_h is now invariant under homomorphism and satisfies (A). That is π_h is an H-property. If one starts from an H-property π, the new property π_h is the same as π; this implies $(\pi_h)_h = \pi_h$.

In some cases π_h is even an R-property. Thus, for example:

LEMMA 7.1. *If π is an isomorphism invariant property satisfying* (A), (C_1) *and* (C_2), *then π_h is an R-property.*

Proof. By Theorem 1.4 and Lemma 2.1, it suffices to show that π_h satisfies also (C_1) and (C_2). To prove (C_1), we consider an ideal A in a ring R such that both A and R/A are π_h-rings. Let R^ϑ be a homomorphic image of R. ϑ induces then a homomorphism of R/A onto R^ϑ/A^ϑ. Since the latter is also a homomorphic image of R and A^ϑ is a homomorphic image of A, it follows by (C_1) and by the definition of π_h that R^ϑ is also a π-ring. This proves that π_h satisfies also (C_1).

To prove (C_2), let $A_1 \subseteq A_2 \subseteq \cdots$ be a non decreasing well ordered sequence of π_h-ideals in R, and let $A = \cup A_\nu$. Let ϑ be a homomorphism of A onto A^ϑ. Then ϑ induces a well ordered sequence of ideals $A_1{}^\vartheta \subseteq A_2{}^\vartheta \subseteq \cdots$ in A^ϑ. Each $A_\nu{}^\vartheta$ is a π-ideal by the definition of π_h. It follows now by (C_2) that, since $\cup A_\nu{}^\vartheta = (\cup A_\nu)^\vartheta = A^\vartheta$, A^ϑ is a π-ring, which proves (C_2).

The following is a method for constructing properties satisfying the conditions of this lema.

To each property π we can define a property of being "π-semi simple" which we shall denote by π.*s.s.* That is: a ring R is a π.*s.s-ring* if it does not contain non zero π-ideals.

LEMMA 7.2. *If π is an H-property satisfying* (D) *then π.s.s. is an isomorphism invariant property which satisfies* (A), (C_1) *and* (C_2).

By definition and by Theorem 1.5 it follows that π.*s.s.* is an isomorphism invariant property which satisfies (A). A proof similar to the proof

of Lemma I. 2. 3, yields (C_1). To prove (C_2), let $A_1 \subseteq A_2 \subseteq \cdots$ be a well ordered sequence of π-semi simple ideals in a ring R. Let $A = \cup A_\nu$. If Q is a π-ideal in A, then since $Q \supseteq Q \cap A_\nu$ and π satisfies (D) it follows that $Q \cap A_\nu$ is a π-ideal. This implies that $Q \cap A_\nu$ is also a π-ideal A_ν, and since A_ν is π-semi simple, we have $Q \cap A_\nu = 0$. Now

$$Q = Q \cap A = Q \cap (\cup A_\nu) = \cup (Q \cap A_\nu) = 0,$$

which proves that A is a $\pi.\boldsymbol{s}.\boldsymbol{s}.$ ring, q. e. d.

It follows now that one can apply Lemma 7. 1 to the property $(\pi.\boldsymbol{s}.\boldsymbol{s}.)_h$, which we denote shortly by $\pi.\boldsymbol{s}.\boldsymbol{h}.$ The definition of this property is the following:

DEFINITION 7. 2. *A ring R is a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring if every homomorphic image of R is π-semi simple.*

Thus by the preceding two lemmas we have:

THEOREM 7. 1. *If π is an H-property satisfying* (D) *then $\pi.\boldsymbol{s}.\boldsymbol{h}.$ is an R-property.*

If we restrict ourselves to associative rings, we can prove more:

THEOREM 7. 2. *If π is a ZH-property of associative rings and satisfies* (D), *then the property $\pi.\boldsymbol{s}.\boldsymbol{h}.$ is an SR-property, i. e. an R-property satisfying* (D) *and* (E). *Furthermore, if π satisfies the condition that R_n is a π-ring if and only if R is a π-ring then $\pi.\boldsymbol{s}.\boldsymbol{h}.$ satisfies also the condition* (G_1), (G_2) *for the radical of the matrix ring.*

Proof. First, we note that under the conditions of the theorem, π-semi simple rings cannot contain nilpotent ideals.

Let R be a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring and let Q be an ideal in R. Suppose Q is not a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring; then it contains an ideal P such that Q/P is not π-semi simple. From the condition of the theorem we can readily deduce that Q/QPQ is also not π-semi simple. Let T/QPQ be a non zero π-ideal in Q/QPQ. It follows by (D) that QTQ/QPQ is also a π-ring; but since R is a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring, R/QPQ must be π-semi simple and, therefore $QTQ \subseteq QPQ$. Consequently, the set $\{x \mid x \in Q, QxQ \subseteq QPQ\}$ constitutes a non zero ideal in R (containing T) which is nilpotent modulo QPQ. This is impossible since in the present case we know that $\pi.\boldsymbol{s}.\boldsymbol{s}.$-rings cannot contain non zero nilpotent ideals. Hence we obtain the validity of (D). The rest of the first part of the theorem follows now from Theorem 5. 1.

Before proceeding with the proof of the second part, we show that if R is a π-semi simple ring then so is R_n. Indeed, let Q be a π-ideal in R_n; then $R_nQR_n \subseteq Q$ is by (D) a π-ideal. Now $R_nQR_n = T_n$ where T is the ideal in R generated by the elements: $rqs, r, s, \varepsilon R$ and where q ranges over all the elements of R which appear in the matrices of Q. By the condition stated in the second part of the theorem it follows that T is a π-ideal, hence $T = 0$. The set $\{x \mid x \varepsilon R, RxR = 0\}$ is a nilpotent ideal in R; it follows, therefore, that this ideal is zero. Consequently, $Q = 0$ and our assertion is proved. We turn now to the proof of (G_1): let R be a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring and suppose R_n is not a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring. Let Q be an ideal in R_n such that R_n/Q is not a π-semi simple ring. By (Z) it follows that R_n/R_nQR_n is also not π-semi simple. The latter is isomorphic with $(R/T)_n$ and R/T is by assumption π-semi simple, hence our assertion implies that $(R/T)_n$ is also a $\pi.\boldsymbol{s}.\boldsymbol{s}.$-ring. Contradiction! To prove (G_2), we note first that if R_n is a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring then R must also be a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring. Indeed, for every ideal T in R, $(R/T)_n \cong R_n/T_n$ and if R/T were to contain a non zero π-ideal then, by the condition of the theorem, the latter would also contain a non zero π-ideal which is impossible. Now let R be a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-semi simple ring and suppose R_n is not such a ring. Let Q be an ideal in R_n which is a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring. By (D) it follows that $R_nQR_n = T_n(\subseteq Q)$ is also a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring. This implies, by the preceding argument, that T is also a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring; hence $T = 0$. But if this is the case, $Q^3 = 0$; but then Q cannot be a $\pi.\boldsymbol{s}.\boldsymbol{s}.$-ring and, evidently, not a $\pi.\boldsymbol{s}.\boldsymbol{h}.$-ring.

The preceding theorem can be applied to many ring properties, and in particular; quasi regularity, nillity and semi nilpotency. The respective radicals will be dealt with in part III.

A by-result of Lemma 7. 2 is the following:

THEOREM 7. 3. *If π is an H-property which satisfies* (D), *then every ring R contains maximal π-semi simple ideals P and each of them is a π-semi simple radical in the sense of section* 1.

The proof follows immediately, since the relation between ideals determined (as defined in section 1) by the property of "π-semi simplicity" satisfies the requirements of Corollary I. 2. 1.

We can prove more:

THEOREM 7. 4. *If π is an R-property which satisfies* (F_s) *(or* (F_r)*), then every ring R contains maximal π-semi simple subrings (right ideals), and $\pi(R) \cap S = 0$ for every such maximal subring (right ideal S).*

The proof of this theorem is similar to the proof of Lemma 7. 2. By proving that if $S_1 \subseteq S_2 \subseteq \cdots$ is a well ordered sequence of π-semi simple subrings of R, then $\cup S_\nu$ is also π-semi simple. This implies the existence of maximal π.s.s.-subrings S. It follows by (D) that $\pi(R) \cap S \subset \pi(R)$ is a π-ideal in $\pi(R)$ and, therefore, in S; we obtain $\pi(R) \cap S = 0$, q. e. d.

In the next section, we turn to another useful method of constructing new radical properties, a method which will be seen to yield a generalization of Baer's Lower Radical ([5]).

8. Additive properties. Many H-properties of rings, like nilpotency, are not R-properties, yet their upper radicals behave very similarly to radicals of R-properties. Those are the H-properties which satisfy the condition of additivity. Namely:

(Add) *If A, B are two π-ideals in a ring R then their union (A, B) is also a π-ideal.*

Such properties will be called A-properties. Evidently, R-properties satisfy this condition. Furthermore, this implies that the union of any *finite* number of π-ideals is a π-ideal.

With the aid of a property π, we define a property π^* as follows:

DEFINITION 8. 1. *A ring R is a π^*-ring if every non zero homomorphic image of R contains non zero π-ideals.*

An equivalent definition is given by the following lemma:

LEMMA 8. 1. *R is a π^*-ring if and only if $R = U(R)$, where $U(R)$ is the upper π-radical.*

Indeed if $R \supset U(R)$, $R/U(R)$ is a non zero π-semi simple homomorphic image of R; hence R cannot be a π^*-ring. Conversely, if R is not a π^*-ring then R properly contains an ideal P such that R/P is π-semi simple. Hence, by Theorem I. 1. 2, $R \supset P \supseteq U(R)$, and the proof is completed.

An associative ring R was called by Levitzki [9] an L-ring if R coincides with its Lower Radical (in the sense of Baer [4]). Noting that Baer's Lower Radical is an uppper radical (in the sense of the present paper), of the property of nilpotency, the preceding lemma shows that the notion of π^*-rings is a generalization of the notion of L-rings.

---

[9] J. Levitzki, "A theorem on polynomial identities," *Proceedings of the American Mathematical Society*, vol. 1 (1950), p. 335.

THEOREM 8.1. *If π is an H-property which satisfies* (E), *then π^* is an R-property which satisfies the same condition; and the π^*-radical coincides with the upper π-radical.*

Proof. Evidently, it follows from the definition of π^* that it is an H-property. First we assert that π^*-semi simplicity is equivalent to π-semi simplicity. Indeed, if R is π-semi simple, then it follows by (E) that $U(P) = 0$ for every ideal P in R. In view of Lemma 7.1, this implies that R does not contain non zero π^*-ideals, i. e., R is π^*-semi simple. The converse is evidently true since π-ideals are also π^*-ideals. Consequently, the validity of (E) for π implies that π^* also satisfies (E). Another result obtained by our assertion is that if R is not a π^*-ring, $R/U(R)$ is a non zero π^*-semi simple image of R. This proves that π^* satisfies (C), hence π^* is an R-property. By Theorem 1.1 it follows readily that $U(R) \supseteq \pi^*(R)$. Since, by Corollary 2.1, $U(R)$ is a π^*-ideal, it follows by Theorem 1.1 that $\pi^*(R) \supseteq U(R)$; hence $\pi^*(R) = U(R)$, q. e. d.

A similar theorem about condition (D) is not true in the general case. But for properties which satisfy the condition of additivity we can show:

THEOREM 8.2. *If π is an A-property which satisfies* (D) *and* (E), *then π^* is an SR-property. Furthermore, if π satisfies also* (D_r) *or* (D_s) *then π^* satisfies also the same condition.*

Proof. From the preceding theorem we see that it remains to prove the validity of (D). We shall prove the second part of the theorem for (D_s) and the proof of (D) and (D_r) is similar.

Let R be a π^*-ring and let S be any subring of R. The proof will follow by showing that $U(S) \supseteq S \cap U(R)$, since then $U(S) = S$ if $U(R) = R$. This is carried out by proving inductively that $U(S) \supseteq S \cap U_\lambda(R)$ for every λ. For $\lambda = 0$ it is evident. If λ is a limit ordinal, then

$$S \cap U_\lambda(R) = S \cap (\bigcup_{\nu<\lambda} U_\nu(R)) = \bigcup_{\nu<\lambda} (S \cap U_\nu(R)) \subseteq U(S).$$

Let λ not be a limit ordinal. Put $V = U_{\lambda-1}(R)$. By definition, $U_\lambda(R)$ is the union of all the ideals P such that P/V is a π-ideal; hence, if $a \in U_\lambda(R) \cap S$, then $a \in P_1 \cup P_2 \cup \cdots \cup P_n$, for a finite number of ideals $\{P_i\}$ such that P_i/V are π-ideals. By the condition (Add), $P = \cup P_i$ is also a π-ideal over V. The quotient $(P \cap S) \cup V/V$ is a subring of P/V, hence by (D_s) it is a π-ring. Since $P \supseteq V$,

$$V \cup (P \cap S)/V \cong P \cap S/[P \cap S \cap V] = P \cap S/V \cap S.$$

Thus, the latter is also a π-ring. By Theorem 1. 7, and from the fact that $U(S) \supset S \cap V$, it follows readily, that since $P \cap S/V \cap S$ is a π-ideal, $U(S) \supset P \cap S$. Thus $a \varepsilon U(S)$ for every $a \varepsilon U_\lambda \cap S$. Consequently $U(S) \supseteq U_\lambda \cap S$, q. e. d.

The proof of (D) (similarly of (D_r)) follows in the same way by replacing S by an ideal (right ideal) B of R and noticing that $V \cup (P \cap B)$ is an ideal (right ideal) in P.

In particular we have by Theorem 2. 3:

COROLLARY 8. 1. *If π is an A-property, then $U(A) = A \cap U(R)$ for every ideal A in R.*

We conclude this seection with showing that:

THEOREM 8. 3. *If π is an H-property which satisfies* (C_1) *then π satisfies the condition* (Add).

Indeed, let A and B be two π-ideals in R. Since $A \cup B/B \cong A/A \cap B$ and the latter is homomorphic with the π-ring A, $A \cup B/B$ is also a π-ring. B is a π-ideal in $A \cup B$ as well as in R; it follows, therefore, by (C_1) that $A \cup B$ is also a π-ring, which proves (Add).

9. Generalization. A close survery of the proofs of the preceding sections (except section 6) shows that no use whatsoever has been made of the operators of the rings and, except for one point (on A-properties), even the fact that rings and ideals contain elements. The whole theory can be developed in a far wider class of mathematical objects. The largest field in which this can be done is that of the Lattice-ordered Bicategories of S. MacLane defined in [7], which satisfy some additional axioms so that the two main isomorphisms hold and some minor properties of ideals in rings.

We do not intend to give here the list of axioms such a category has to satisfy, but only a list of conditions, some of which may be considered as axioms, others as lemmas, to be valid in such a category in order that the whole theory can be developed in it. It is worth noting that the whole theory of radicals is just a relation between injections and projections of a bicategory. The proof in the general case can be obtained by a simple change of 'objects' and 'normal objects,' 'supermaps' instead of 'rings,' 'ideals' and 'homomorphisms.'

In giving the list of conditions we shall make use of notions and theorems given in [7], with the only difference of using the zero symbol 0_A instead of the multiplicative symbol 1_A used in [7].

Let $\mathfrak{C}$ be a lattice-ordered bicategory satisfying the axioms of part II of [7] (pp. 495-507). Such a category contains objects denoted by $A, B, \cdots, R, S, \cdots$ and mappings $\alpha, \beta, \gamma, \cdots$. Recall the following facts of [7]: The set of all subobjects $A \subset F$ form a complete lattice $\mathfrak{S}(R)$ with the zero 0_R, and R as zero element.

If $\alpha: R \to R'$, then α_s denotes in [6] the mapping of $\mathfrak{S}(R)$ induced by α (for details see [7], p. 500). We shall use the notation α instead of α_s and put αA for the image $\alpha_s A$.

The kernel K_α of α is by definition the l. u. b.$\{A \mid A \subset R, \alpha A = 0_R\}$. The set of all kernels of R is exactly the set of all subobjects of R *normal* in R (in the sense of [7]). With each normal object A of R, there is associated a unique projection $i_A: R \to R/A$, with the unique quotient object R/A.

A simple observation shows that the set $\mathfrak{N}(R)$ of all normal subobjects of R constitutes a complete lattice. This is not necessarily a sublattice of $\mathfrak{S}(R)$. It is true that the intersection of normal subobjects in $\mathfrak{N}(R)$ is the same intersection as in $\mathfrak{S}(R)$ ([7], p. 507), but the union $\cup N_\nu$ in $\mathfrak{N}(R)$ is defined as the intersection of all normal subobjects of R containing all N_ν. This is the general set-up in which one can state the results of the preceding theory. The objects R will replace the rings and the kernels (= normal subobjects) will replace the ideals. In order for the proofs to hold, one has to assume that the complete lattice satisfies some additional requirements.

I) If $\alpha: R \to \alpha R$ is a supermap (we call it a homomorphism (onto) to make it similar to the language used here), then α maps normal subobjects of R onto normal subobjects of αR, and $\alpha(\cup N_i) = \cup (\alpha N_i)$.

II) If $\alpha: R \to \alpha R$ is a homomorphism, and $\alpha^{-1}A$ denotes the inverse image of a normal subobject A of αR ([7], p. 507) then α^{-1} [10] is a lattice isomorphism between the lattice of all normal subobjects of R lying in the interval $K_\alpha \subset R$ (K_α the kernel of α) and the lattice $\mathfrak{N}(\alpha R)$; and $R/\alpha^{-1}B \leftrightarrow \alpha R/B$ for every $B \in \mathfrak{N}(\alpha R)$.

III) If $R \to \alpha R$ is a homomorphism, $A \in \mathfrak{N}(R)$, then α induces a homomorphism $R/A \to \alpha R/\alpha A$.

IV) $Q, A, B, \in \mathfrak{N}(R)$. If $R \supset A \supset B$ then $B \in \mathfrak{N}(A)$,[11] and so $Q \cap A \in \mathfrak{N}(Q)$.

---

[10] Denoted by α^*, in [7].

[11] This fact was pointed out in [7], p. 507.

V) If $R \supset B \supset A, A, B \in \mathfrak{R}(R)$, then there is a homomorphism: $R/A \to R/B$ such that for every $C \supset A$, $C \in \mathfrak{R}(R)$, C/A exists, and its image is $(C \cup B)/B$. Furthermore, by (II) we have: $(R/A)/(B/A) \leftrightarrow R/B$ (The Second Isomorphism theorem).

VI) For $A, B \in \mathfrak{R}(R)$, there exists an equivalence (isomorphism) $A \cup B/A \leftrightarrow B/A \cap B$ (The First Isomorphism theorem).

VII) If $A_1 \subset A_2 \subset \cdots$ is a well ordered non decreasing sequence in $\mathfrak{R}(R)$ then $\cup A_i$ in $\mathfrak{R}(R)$ is the same $\cup A_i$ in $\mathfrak{R}(B)$ for every normal object $B \supset A_i$.

These conditions are fundamental for the theory developed in sections 1, 2, 7, 8 (except Theorem 2.1 and its corollary). The second part of condition II was used only at Theorem 2.3. For the sake of Lemma 7.2 we need also:

VIII) If $A_1 \subset A_2 \subset \cdots$ is of the preceding type, then $Q \cap (\cup A_\nu) = \cup (Q \cap A_\nu)$ for any $Q \in \mathfrak{R}(R)$; and for the sake of Theorem 8.2, we need (VIII) and:

IX) If $\{P_\alpha\}$ is a set of $\mathfrak{R}(R)$, $B \in \mathfrak{R}(R)$, $A \in \mathfrak{R}(B)$ and such that $A \supset (\cup P_i) \cap B$ for every finite set $P_1, \cdots, P_k$ chosen out of $\{P_\alpha\}$ then $A \supset P \cap B$, where P is the union of *all* P_α.

In the proofs of sections 3 and 8, and of Theorem 2.1 and its corollary, one notices the resemblance between the problem of one sided ideals and subrings. These two cases are particular applications of similar problems which arise for the bicategories of the preceding type. The general situation is as follows: we assume that to each object R belongs a complete lattice of subobjects $\mathfrak{J}(R)$ (right ideals, subrings). These subobjects should satisfy conditions of a similar nature to the conditions satisfied by $\mathfrak{R}(R)$. Namely:

I′) If $\alpha: R \to \alpha R$ is a homomorphism, then α induces a lattice homomorphism of $\mathfrak{J}(R)$ onto $\mathfrak{J}(\alpha R)$ such that $\alpha(\cup J_i) = \cup (\alpha J_i)$. Furthermore, $J_i \to \alpha J_i$ is a homomorphism.

II′) The lattice of normal subobjects $\mathfrak{R}(R)$ is a sublattice of $\mathfrak{J}(R)$.

III′) If $A \in \mathfrak{R}(R)$, $J \in \mathfrak{J}(R)$, then $A \cap J \in \mathfrak{R}(J)$ and $\mathfrak{J}(A)$, and there is an isomorphism $A \cup J/A \leftrightarrow J/A \cap J$.

IV′) If $\{A_\alpha\}$ is a set of $\mathfrak{R}(R)\,(J(R))$ and $A_\alpha \subset J \in J(R)$, then $A_\alpha \in \mathfrak{R}(J)$ and $\cup A_\alpha$ in $\mathfrak{R}(R)\,(\mathfrak{J}(R))$ is the same object as $\cup A_\alpha$ in $\mathfrak{R}(J)\,(\mathfrak{J}(J))$.

V') If $R \supset J \supset A$, $A \in \mathfrak{N}(R)$, $J \in \mathfrak{J}(R)$, then J/A exists and belongs to $\mathfrak{J}(R/A)$.

And for the sake of Theorem 8. 2, we need the previous conditions (VIII) and (IX) to hold also for $B, Q \in \mathfrak{J}(R)$ and the fact that if $Q \supset J$, $Q, J \in \mathfrak{J}(R)$ then $J \in \mathfrak{J}(Q)$.

Even without going as far as bicategories one readily observes that the whole theory can be developed without any additional efforts for groups, semi groups, loops, etc., or, more generally, to abstract algebras in the sense of Birkhoff (e. g. [2]), where ideals should be defined as kernels of homomorphisms. These abstract algebras should be of a more restrictive type than those studied in [2], Ch. VI (also, Goldie [6]). Namely: one has to assume that the operations of the abstract algebras $R = (R, \Omega)$ are finitary functions $f_\alpha(x_1, \cdots, x_n)$, $n < \infty$; $f_\alpha \in \Omega$ and that:

(A) These algebras contain a selected one element, denoted by zero.

(B) The congruent relations of the algebras are uniquely determined by the set of elements which are congruent to zero (i. e. by the kernel = ideal).

(C) All congruent relations considered must be permutable (in the sense of [2], Ch. VI).

By the results of [2], Ch. VI (see also [6]), one readily shows that these algebras satisfy all conditions imposed on the bicategories.. In particular we remark that if $A_1 \subseteq A_2 \subseteq \cdots$ is a well ordered sequence of ideals in such an algebra R then $\cup A_\nu$ is just the set-theoretic union, which proves (VII), (VIII).

To these algebras one can extend part of the theory of radicals as developed by Brown-McCoy in [3] and [4]. Their whole theory can be extended to algebras of the type mentioned above which are loops with respect to one of their binary functions: $f_1(x, y) = x + y$. In the next section we develop their theory first, partly, in the general case and then the whole theory for loop algebras.

9. Brown-McCoy's theory of *F*-radical. Furthermore, we assume, in this section, that to every algebra considered there corresponds a complete lattice $\mathfrak{R}$ of subsets containing the zero. By a $\mathfrak{R}$-*subset* of R we shall mean a subset of R which belongs to $\mathfrak{R}$. We also assume that for every homomorphism ϑ of R the $\mathfrak{R}$-subsets of R^ϑ are the images of the $\mathfrak{R}$-subsets of R and the inverse images of the first is a subset of the latter.

Let $a \to F(a)$ be a mapping of elements of algebras R onto $\mathfrak{R}$-subsets of R. This mapping is assumed to be homomorphic invariant; that is: $F(a^\vartheta) = F^\vartheta(a)$.[12] Following [3], we define: a $\mathfrak{R}$-subset T of R is called *maximal modular* if there exists an element $a \in R$ such that $F(a) \subseteq T$ and T is a maximal $\mathfrak{R}$-subset of R which does not contain a. An ideal A in R is said to be *F-primitive* if A is a maximal ideal contained in some maximal modular $\mathfrak{R}$-subset of R. An algebra R is *F-primitive* if the zero is a primitive ideal in R. An ideal P is called *F-regular* if $a \in F(a)$ for every $a \in A$.

It is readily seen that F-regularity is an HI-property in the sense of section 1, hence it yields an upper F-radical which we denote by $F(R)$.

THEOREM 9.1. *If the primitive ideals P are $\bar{F}$-ideals,*[13] *then $F(R)$ is the F-radical of R and it is the intersection of all primitive ideals of R.*

Proof. Let $A \supset B$ be two ideals in R such that A/B is F-regular in R/B. We wish to show the existence of a primitive ideal $P \supseteq B$ such that $A \supset A \cap P$. Without loss of generality we assume that $B = 0$. Since A is not F-regular, $a \notin F(a)$ for some $a \in A$. By Zorn's Maximum Principle we deduce the existence of a maximal modular $\mathfrak{R}$-subset S of R containing $F(a)$ and excluding a.[14] Let P be a primitive ideal contained in S, then since $a \notin S \cap A$, $A \supset A \cap P$, q. e. d. Our theorem follows immediately as a consequence of Theorem I. 2. 4.

Generally, it cannot be shown that F-primitive ideals are $\bar{F}$-ideals. Furthermore, it is not certain that if A is F-primitive in R then R/A is an F-primitive algebra. The following are some special cases for which this is true.

THEOREM 9.2. *If the sets $F(a)$ (the set $\mathfrak{R}$) are the set of all ideals, then if A is an F-primitive ideal in R, R/A is F-primitive and every primitive algebra is F-semi simple.*

Proof. In the present case F-primitivity coincides with maximal modularity, hence if A is F-primitive $F(a) \subseteq A$ for some $a \notin A$. In the quotient R/A, the image $\bar{a}$ of the element is a non zero element such that $F(\bar{a}) = 0$. By the maximality of A with respect to A, one readily observes that every non

---

[12] Here $F\vartheta(a) = \{x \mid x \in R, x = y^\vartheta \text{ for some } y \in F(a)\}$.

[13] Meaning that R/P does not contain non zero F-regular ideals.

[14] Here we have to assume that the union of a linearly ordered set of $\mathfrak{R}$-subsets is a $\mathfrak{R}$-subset which is their set-theoretic union. This holds for example if the $\mathfrak{R}$-subsets are all subalgebras.

zero ideal in the quotient R/A must contain $\bar{a}$ and thus cannot be F-regular. This proves that R/A is primitive as well as F-semi simple.

The preceding two theorems present a generalization of the F-radicals as developed in [4] and in example 3 of [3], section 6. The incorporation of the theory developed in [3] is achieved by restricting ourselves to abstract algebras which are loops with regard to one of their functions $f(a, b) = a + b$ and assuming that the $\mathfrak{R}$-subsets are subloops. We refer to such algebras as loop-algebras.

Now let $a \to F(a)$ be a homomorphic invariant mapping of the elements of loop-algebras onto $\mathfrak{R}$-subsets which satisfies the following conditions of [3]:

P. $F(a + b) \subseteq F(a) \cup (b)_1$.

where $F(a) \cup (b)_1$ denotes the minimal loop-subalgebra of R which contains both $F(a)$ and the element b.

In a similar way to the proof of the lemma of [3, section 8], one verifies that if A is F-primitive in R then R/A is an F-primitive ring. Hence the theory developed in [3] section 8 is now a consequence of Theorem 9. 1 and the following lemma:

LEMMA 9. 1. *If R is an F-primitive loop-algebra then R is F-semi simple.*

Indeed, let R be primitive with respect to a and the maximal modular $\mathfrak{R}$-subloop S. Let P be a non zero ideal in R. The natural homomorphism of R onto R/P maps S onto the $\mathfrak{R}$-subloop $P(S)/P$ of R/P. Since $P \not\subseteq S$, $P(S) \supset S$. Hence $a \in P(S)$. Thus $a \equiv s \pmod{P}$ for some $s \in S$ and, therefore, $a - s = p \in P$. It follows now by condition (P) that

$$F(p) \subseteq F(a) \cup (s)_1 \subseteq F(a) \cup S = S.$$

Assuming that P is F-regular, we obtain $a - s = p \in F(p) \subseteq S$ which implies $a = p + s \in S$. Contradiction.

Next we assume that the mapping $a \to F(a)$ is homomorphic invariant and satisfies the second condition of [3]:

(P_2) If $b \in F(a)$ then $F(a + b) \subseteq F(a)$.

This condition does not imply Theorem 9. 1, yet we are able to show that in this case:

THEOREM 9. 3. *If $a \to F(a)$ is a homomorphic invariant and satisfies (P_2) then the property F-regularity is an RI-property.*

In view of Theorem 2 of [3] and of the results of section 1, this theorem yields the results of sections 1-3 of [3].

Proof. We show the validity of (C_1) and (C_2), and the proof of the theorem will follow by Theorem 1. 4. The validity of (C_2) is evident. To prove (C_1) we consider the ideals $A \supseteq B$ of R such that A/B and B are F-regular. The natural homomorphism of R onto R/B maps $F(a)$ onto $B \cup F(a)/B$. Hence for every $a \in A$, $\bar{a} = (a + B/B) \in F(\bar{a}) = F(a) \cup B/B$, i. e., $a = x + b$ for some $x \in F(a)$. Thus, $a - x = b \in B$. It follows now by (P_2) that $F(b) \subseteq F(a)$. Since B is F-regular, $b = a - x \in F(b) \subseteq F(a)$ and, therefore, $a \in F(a)$. This proves that A is F-regular,, q. e. d.

HEBREW UNIVERSITY
AND THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] S. A. Amitsur, "A general theory for radicals, I. Radicals in complete lattices," *American Journal of Mathematics*, vol. 74 (1952), pp. 774-786.

[2] G. Birkhoff, "Lattice theory," *American Mathematical Society Colloquium Publications*, vol. xxv, 1948.

[3] B. Brown and N. H. McCoy, "Some theorems on groups with applications to ring theory," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 302-311.

[4] ———, "Radicals and subdirect sums," *American Journal of Mathematics*, vol. 69 (1947), pp. 46-58.

[5] R. Baer, "Radical ideals," *American Journal of Mathematics*, vol. 69 (1947), pp. 46-58.

[6] A. W. Goldie, "The Jordan-Hölder theorem for general abstract algebras," *Proceedings of the London Mathematical Society*, vol. 52 (1950), ser. 2, pp. 107-113.

[7] S. MacLane, "Duality for groups," *Bulletin of American Mathematical Society*, vol. 56 (1950), pp. 485-516.

A GENERAL THEORY OF RADICALS.*

III. Applications.

By S. A. Amitsur.

The present paper contains applications of the theory developed in the first two parts ([1], [2]) to theory of rings, in particular to associative rings. The results of the preceding parts [1], and [2] will be quoted by the numbers I, or II, followed by the number of the result quoted.

In section 1 we outline an extension of the theory of nilpotent ideals in associative rings to non associative (and non distributive) systems, obtaining a generalization of Baer's Lowe Radical ([3]) and proving that this generalized radical is also the intersection of all prime ideals, a result which was proved by Levitzki ([8]). Continuing the study of nilpotent ideals, we restrict ourselves in section 2 to associative rings. In this section, some new results on the Lower Radical are obtained as well as some short cuts in the proofs and new results for the semi-nilpotent radical ([7]). In view of the results obtained in section 2, it seems that Baer's Lower Radical is the best 'minimal' extension of the nilpotent radical of rings satisfying the minimum condition for right or left ideals. In section 3 we approach some of the known radicals from the point of view of the present theory. Thus we obtain alternative proofs for: Jacobson's radical ([6]), the locally finite kernel ([10]) and McCoy's radical ([11]). In the last section we utilize the theory for obtaining new radicals, among which is an FI-radical. A ring R is an FI-ring if every homomorphic image R' of R is an I-ring ([10], i. e. every non nil right ideal in R' contains an idempotent). It is shown that this property FI yields a radical which satisfies all the ordinary requirements a radical is supposed to satisfy.

1. Nilpotency and solvability. In the present section we consider rings (semi groups) R not necessarily associative (or distributive) and we outline an extension of the theory of nilpotent ideals to the non-associative case. The whole development here will hold also, without any changes in the proof, in general abstract algebras of the type considered in the last section of II, which are semi groups with respect to a binary operation $f(a, b)$.

* Received September 10, 1952; substantially revised March 7, 1953.

Let Q_1, Q_2 be two subsets of such a ring R. We denote by Q_1Q_2 the set of all element q_1q_2, $q_1 \in Q_1$, $q_2 \in Q_2$. In particular we put $Q' = QQ$ and inductively $Q^{(n)} = (Q^{(n-1)})'$. Note that $(Q^{(n)})^{(m)} = Q^{(nm)}$. An ideal P in R is called *a prime ideal* (in R) if: $P_1P_2 \subseteq P$, where $P_i \subseteq P$ are ideals in R, implies that either $P_1 \subseteq P$ or $P_2 \subseteq P$.

Defining a ring R to be a *ζ-algebra* if $R' = 0$,[1] we obtain an H-property ζ of rings. The upper ζ-radical, which we denote by $\zeta(R)$, will be called in this paper the *Baer's radical*. It is known (Levitzki [8]) that in the case of an associative ring R, Baer's radical is the intersection of all the prime ideals in R (McCoy's radical [11]). It will be shown here that this result is true for non associative ring, semi groups and more generally this is true for the abstract algebras of the type mentioned above. That is:

THEOREM 1.1. *$\zeta(R)$ is the intersection of all the prime ideals of R.*

Proof. Observing that prime ideals are $\overline{\zeta}$-ideals, we deduce from Theorem I. 1. 2 that the intersection X of all the prime ideals of R contains $\zeta(R)$. The theorem will be proved if we show that for every $a \notin \zeta(R)$ there is a prime ideal Q in R which does not contain the element a, since then $X \subseteq \zeta(R)$. Indeed, put $a_1 = a$; let A_1 be the minimal ideal in R containing a_1 and $\zeta(R)$. Since $a_1 \notin \zeta(R)$, $A_1 \supset \zeta(R)$. Hence $A'_1 \not\subseteq \zeta(R)$. Let a_2 be any element of A'_1 which does not belong to $\zeta(R)$. In a similar way we can inductively obtain a sequence of elements $a_1, a_2, \cdots$ such that $a_i \notin \zeta(R)$ and $a_{i+1} \in A'_i$, where A_i is the minimal ideal in R containing $\zeta(R)$ and a_i. By Zorn's lemma there exists a maximal ideal P in R containing $\zeta(R)$ and excluding the sequence $\{a_i\}$. The ideal P is the required ideal. For if $P_1P_2 \subseteq P$ and $P_i \supset P$ then the maximality of P implies that $a_j \in P_1$ and $a_k \in P_2$, for some j and k; hence $P_1 \cap P_2$ contains an element a_l of the sequence $\{a_i\}$. But this implies that $a_{l+1} \in A'_l \subseteq P_1P_2 \subseteq P$ which is a contradiction.[2]

The preceding proof provides an alternative characterization of $\zeta(R)$. We call a sequence $\{q_i\}$ of elements of R an *m-sequence* if $q_{i+1} \in Q'_i$, $i = 1, 2, \cdots$, where Q_i denotes the minimal ideal in R containing q_i. An m-sequence is said to be a *null sequence* if it contains only a finite number of non zero elements.

Define a property μ of ideals in R as follows: an ideal in A in R is a *μ-ideal* in R if every m-sequence of elements of A is a null sequence.

We prove:

---

[1] Recall that P is said to be a $\overline{\zeta}$-ideal in R if the quotient R/P is ζ-semi simple. For associative rings, the notion of $\overline{\zeta}$-ideals coincide with the notion of radical-ideals of [3].

[2] This proof resembles that of the lemma of [9].

THEOREM 1.2. *The property μ is an RI-property (in the sense of II-section 2) and $\mu(R) = \zeta(R)$.*

Let ϑ be a homomorphism of R onto R^ϑ and Q be an ideal in R. Let P be any ideal in R, such $P = Q^\vartheta$. First we show that for any m-sequence $\{b_i\}$ in P, there exists an m-sequence $\{a_i\}$ in Q such that $b_i = a_i^\vartheta$, $i = 1, 2, \cdots$. Choose a_1, any element of Q such that $a_1^\vartheta = b_1$. Suppose that $a_1, \cdots, a_n$ have been so chosen that $a_i \varepsilon Q$, $a_i^\vartheta = b_i$, $i = 1, \cdots, n$ and $a_{i+1} \varepsilon A'_i$ where A_i denotes the minimal ideal in R containing a_i. We determine a_{n+1} as follows: since $a_n^\vartheta = b_n$ it is evident that $A_n^\vartheta = B_n$, where B_n is the minimal ideal in R which contains b_n. Since $(A'_n)^\vartheta = B'_n$ and $b_{n+1} \varepsilon B'_n$ we can find $a_{n+1} \varepsilon A'_n$ such that $a_{n+1}^\vartheta = b_{n+1}$ and evidently $a_{n+1} \varepsilon Q$. This proceduce determines inductively the required sequence.

Consequently, if every m-sequence of the ideal Q is a null sequence then the ideal $P = Q^\vartheta$ has the same property. This proves that μ is an *HI*-property. That μ is an *RI*-property will be shown by proving the validity of II. (C). Let A be not a μ-ideal in R, then A contains a non null sequence $\{a_i\}$. Consider the maximal ideal B contained in A and which does not contain the sequence $\{a_i\}$. Evidently $A \supset B$. If $A \supseteq P \supset B$ then the maximality of B implies that $a_i \varepsilon P$ for some i. Hence $a_j \varepsilon P$ for every $j \geq i$. Thus P/B contains the non null sequence $\bar{a}_i, \bar{a}_{i+1}, \cdots$ which proves that P/B cannot be a μ-ideal, and II. (C) is proved.

To prove $\mu(R) = \zeta(R)$, we note first that a ζ-ideal in R is a μ-ideal. By Corollary I. 1. 5 it remains, therefore, to show that ζ-semi simplicity implies μ-semi simplicity. Indeed, let R be a ζ-semi simple ring and let $A \neq 0$ be an ideal in R. Choose a_1 any non zero element of A; since $A'_1 \neq 0$ take $a_2 \neq 0$ any element of A'_1. Inductively we obtain a non null m-sequence $\{a_i\}$ in A, which shows that R is also μ-semi simple.

Remark 1. 1. The preceding proof, with the exception of the last part, is readily seen to hold, with some slight changes, for the following similar properties of rings (semi groups, etc.). A sequence $\{p_i\}$ in R is a m_0-sequence if $p_{i+1} \varepsilon P_i$ where P_i is the minimal sub ring of R containing p_i (alternative definitions: $P_i = p_i R$, $P_i = Rp_i$, etc.); and R is a μ_0-algeba if every m_0-sequence is a null sequence.

The second aim of the present section is to outline an extension of the theory of nilpotent ideals of associative rings to non associative rings (semi groups, or general multiplicative algebras). It seems that in the general case the property of solvability should replace the property of nilpotency.

We call a subset Q of R a *solvable* subset if $Q^{(n)} = 0$ for some integer n; and a ring R is said to be a *σ-ring* if R is solvable. We call a ring R a

$s\sigma$-ring (semi solvable)[3] if every finite subset of R is solvable. Furthermore, an element $a \in R$ is said to be a nil element if the set (a) containing the element a is solvable, and a ring R is a *nil-ring* (ν-ring) if all its elements are nil elements.

THEOREM 1.3. 1) *σ is an H-property which satisfies* (C) *and* (D_s)[4] *of* II. 2) *$s\sigma$ and ν are R-properties which satisfy* $II(D_s)$ *(and hence* $II(D)$*);* *and* $\nu(R) \supseteq s\sigma(R) \supseteq \sigma(R) \supseteq \zeta(R)$.

Proof. Evidently, each of the properties are H-properties and satisfy II. (F_s). σ satisfies II. (C_1), for if A/B is solvable and B is solvable then $A^{(n)} \subseteq B$ for some n and $B^{(m)} = 0$, and hence $A^{(nm)} = 0$. A similar proof holds for $s\sigma$ and ν. Since these last two properties evidently satisfy II. (C_2), it follows by Theorem II. 1. 4 that they are R-properties. The relation between the various radicals follows by Theorem I. 1. 6 as a consequence of the fact that a ζ-ring is solvable, solvability implies semi solvability and, finitely, a semi solvable ring is a nil ring.

There are many other aspects of extending the theory of nilpotent ideals of associative rings to the non associative case. Thus, for instance, nilpotency may be defined as follows: a subset Q of R is nilpotent if there exists an integer n such that all the possible multiples of n elements of Q are zero. Similarly we can define semi nilpotency and nillity. Each of these properties is readily seen to be an H-property. But the main disadvantage of the nilpotency defined in this manner is that it does not satisfy II. (C_1) and, therefore, its allied properties: semi nilpotency and nillity are not; generally, R-properties and do not yield a radical. The validity of II. (C_1) seems the main reason for preferring solvability in the theory of non associative rings over the other generalisations of nilpotency.

Remark 1. 2. Radicals of a similar type can be defined in abstract algebras of the type considered in section II, but which need not be ring or semi groups. We assume that these algebras possess a function $f(x_1, \cdots, x_n)$, $n < \infty$, (e. g. Lie triple systems) about which we assume that $f(x_1, \cdots, x_n) = 0$ if one of the $x_i = 0$. As in section II ideals will be the kernels of homomorphisms. An ideal will be said to be prime if

$$f(P_1, \cdots, P_n) = \{f(p_1, \cdots, p_n), p_i \in P_i\} \subseteq P,$$

where P_i are ideals, implies that at least one $P_i \subseteq P$. Denoting

$$R' = f(R, \cdots, R) = \{f(r_i, \cdots, r_n), r_i \in R\},$$

we define a ring to be a ζ-ring if $R' = 0$.

---

[3] This is a generalisation of the notion of semi-nilpotency defined in [7].

[4] By Theorem II. 8. 3 it follows that σ satisfies also the additivity condition II. (Add) of section 8.

2. Nilpotency and semi-nilpotency of associative rings and semi-groups. We note that when associativity is assumed the two main properties solvability and nilpotency and, therefore, their issues are equivalent. Another point to notice is the difference between the present definition of semi-solvability and the definition of semi nilpotency given in [7]. Our definition seems to be more restrictive by requiring the solvability of every finite subset, but evidently the two definitions are equivalent if associativity is assumed. In the present section we deal with further properties of the radicals defined in the preceding one. We shall consider only associative rings but we remark that the following proofs and results hold also (with slight modifications) for associative semi-groups.

Since in the present case solvability and semi-solvability are equivalent to nilpotency and semi-nilpotency we shall also refer to the latter by the respective notation σ and $s\sigma$.

In addition to the properties of $s\sigma$ given in Theorem 1. 3 we prove:

THEOREM 2. 1. *$s\sigma$ is an RZ-property*[5] *which satisfies* II. (F_r).

In view of Theorem 1. 3 and the validity of (2), it remains to show the validity of II. (F_r). This will follow by proving Lemma II. 3. 1 by a method due to J. Levitzki (compare with [10] Lemma 7. 3). Let J be a right semi-nilpotent ideal in R. Consider the ideal RJ. Let $t_1, \cdots, t_n$ be any finite set of elements of RJ; then $t_i = \sum_k r_{ik}s_{ik}$, $r_{ik} \in R$, $s_{ik} \in J$. Denote by $[t]$, $[r]$ and $[s]$ the module generated by the elements $\{t_i\}$, $\{r_{ik}\}$, and $\{s_{ik}\}$ respectively, then $[t] \subseteq [r][s]$. Hence $[t]^{m+1} \subseteq [r]([s][r])^m[s]$. The module $[s][r] \subseteq [sr]$ and the latter is generated by a finite number of elements of J. Hence $([s][r])^m = 0$ for some m. This proves that $[t]^{m+1} = 0$. Thus RJ is semi nilpotent.

The known properties of the semi nilpotent radical ([7]) follow now immediately by the results of part II. Additional results follow by Theorem II. 2. 2 and II. 6. 3 namely:

COROLLARY 2. 1. 1) $s\sigma(A) = s\sigma(R) \cap A$ *for every ideal A in the ring R.* 2) $s\sigma(R_n) = s\sigma(R)_n$.

Next we turn to a further study of the Baer's radical $\zeta(R)$ of associative rings.

For associative rings, ζ-semi simplicity implies σ-semi simplicity and hence by Corollary II. 1. 6 and by Theorem 1. 3 it follows that $\sigma(R) = \zeta(R)$. We have already seen that σ is an HZ-property which satisfies II. (C_1) and

[5] For definition see section II. 6.

II. (D_s) (and evidently II. (D_r) and II. (D)). Furthermore σ satisfies II. (E): this follows from the fact that if P is a nilpotent ideal (a σ-ideal!) in an ideal Q of a σ-semi simple ring R, then since $(RPR)^{2n+1} \subseteq RP^nR$, RPR is a nilpotent ideal in R, which implies $RPR = 0$ and consequently $P = 0$. It follows now by Theorem II. 8. 2 that σ^* is an RZ-property which satisfies II. (D_s) and II. (D_r). A ring R was called an L-ring[6] an L-ring if $R = \zeta(R) = \sigma(R)$.

In our notation this is equivalent to the definition of a σ^*-ring. Using the notation of L-rings we obtain, in view of the definition of σ^* and of Lemma II. 8. 1:

COROLLARY 2. 2. *A ring R is an L-ring if and only if every non zero homomorphic image of R contain non zero nilpotent ideals.*

The validity of (D_s) implies:

COROLLARY 2. 3. *Every subring of an L-ring is an L-ring.*

We wish now to show that the property L ($= \sigma^*$) satisfies II. (F_r). Let R be an L-semi simple ring and let J be a right ideal in R such that $L(J) = J$. The annihilator of J in J is the maximal nilpotent ideal in J. For, if Q is a nilpotent ideal in J, then QJ is a right nilpotent ideal in R. The L-semi simplicity of R readily implies that R does not contain non zero right nilpotent ideals and hence $QJ = 0$. Thus $Q \subseteq N$, where N is the left annihilator of J in J. By Theorem II. 1. 1 and by the definition of the upper radical, it is readily seen that $L(J) = N$. Hence $J^2 = 0$. By the L-semi simplicity it follows now that $J = 0$.

Thus, we obtain by Theorem II. 2. 2 and by Theorem II. 6. 3:

COROLLARY 2. 4. 1) $L(A) = L(R) \cap A$ *for every ideal A in a ring R.* 2) $L(R_n) = L(R)_n$.

We conclude with a remark that the preceding results with the exception of the second parts of the Corollaries 2. 1 and 2. 4 hold also for associative semi groups.

3. Various applications. We bring now some alternative proofs of some of the known radicals, in particular those which do not enter in the class of F-radicals, discussed in [4].

A. *Jacobson's radical.* This radical and its generalizations were dealt with from the point of view of F-radicals by several authors. We shall

[6] This is the notation used by J. Levitzki in: "A theorem on polynomial identities," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 334-341.

deal here only with the associative case. Recall that a ring R was called a quasi-regular ring (QR-ring) if for every $r \in R$ there exists an $s \in R$ such that $r + s + rs = 0$.[7] This property is readily seen to be an HZ-property and satisfies II. (D) and II. (D_r). Furthermore, QR satisfies the condition of Theorem II. 6. 6. Indeed, if R is not a QR-ring, let $a \in R$ be not a quasi regular element. Consider the maximal right ideal J containing the ideal $\{x + ax\}$ and excluding a. By the proof of Theorem 17 of [6] it follows that $R/(J:R)$ does not contain right QR-ideals and this proves Theorem II. 6. 6. Hence QR is an RZ-property and satisfies also II. (F_r). The validity of II. (E) follows by Theorem II. 5. 1. Thus most of the properties of the QR-radical obtained in [6] are deduced from the results obtained here.

B. *The locally finite kernel.* An associative algebra R over a field F is said to be *locally finite* (LF-ring) of every finitely generated subalgebra of R is finite over F. Levitzki has recently shown in [10] the existence of the LF-radical, which was called the locally finite kernel. The present theory provides a short cut in Levitzki's proofs and yields further properties of this radical.

The property LF is readily seen to be an HZ-property which satisfies II. (C_2) and II. (D_s) and evidently II. (D). The validity of II. (C_1) follows by [10] Lemma 7. 4, and II. (F_r) follows by [10] Lemma 7. 3. Since LF is also antiisomorphism invariant, a similar condition holds also for left ideals (Remark 1. 1). The validity of all these conditions yields the results on the locally finite kernel obtained in [10]. Additional results now obtained are: $LF(A) = LF(R) \cap A$ for every ideal in an algebra R; and $LF(R)_n = LF(R_n)$.

C. *McCoy's radical.* This radical, which was defined in [11], was proved to be equal to Bear's radical. It can be also approached from the present point of view of HI-properties. Calling an ideal A in a ring an M-ideal if every m-system of A contains the zero, we obtain an HI-property in the sense of section II. 1. The prime ideals are readily seen to be $\bar{M}$-ideals, and if $A \supset B$ are ideals in R such that A/B is not an M-ideal, one can find a prime ideal Q[8] in R such that $A \supset A \cap Q \supseteq B$. It follows now by Theorem I, 2. 4 that M is an RI-property and that the M-radical is the intersection of all the prime ideals.

---

[7] Usually, quasi-regularity is defined in a ring but it is known that an ideal A (right, left or two sided) in a ring R is quasi regular in R if and only if A is a quasi-regular ring.

[8] The ideal Q can be chosen as a maximal ideal $Q \supset B$ such that Q/B excludes a certain m-system of A/B.

4. New radicals. A. *PI-radical.* For the present radical we consider only associative algebras over a field F. An algebra R is said to be a *PI*-ring if it satisfies a polynomial identity $f(x_1, \cdots, x_n) = 0$ where the coefficients of f belong to F. This property is readily seen to be an *HZ*-property which satisfies II. (D_s) and hence also II. (D). Furthermore, *PI* satisfies II. (C_1) and II.(E). Indeed, if R possesses an ideal A such that R/A satisfies the identity $g(x_1, \cdots, x_n) = 0$ and such that A satisfies the identity $f(x_1, \cdots, x_n) = 0$, then R satisfies the identity

$$f[g(x_{11}, \cdots, x_{1n}), \cdots, g(x_{m1}, \cdots, x_{mn})] = 0$$

in the nm indeterminates x_{ik}. This proves II. (C_1). To prove II. (E), let R be a *PI*-semi simple ring and let A be an ideal in R. If Q is an ideal in A which satisfies an identity, then $AQA \subseteq Q$ is a *PI*-ideal in R, hence $AQA = 0$. But the totality of the elements $x \in A$ such that $AxA = 0$ constitute a nilpotent ideal in R, hence, by II. (Z), it must be zero; and thus II. (E) is proved.

It follows now by Theorem II. 8. 3 that *PI* is an additive-property and, therefore, *PI** is an *RZ*-property and satisfies II. (E) and (D_s). We recall that a ring R is a *PI**-ring if every non zero homomorphic image of R contains non zero ideals which satisfy a polynomial identity.

The property *PI** is an example of an R-property which does not satisfy II. (F_r). For, the ring R of all finite matrices is a simple ring and it is known that this ring does not satisfy a polynomial identity, yet the right ideal $c_{11}R$, i. e., the ring of all one rowed matrices satisfies the identity $(x_1x_2 - x_2x_1)x_3 = 0$.

B. ***π.s.h.-radicals.*** The methods of section II. 7 can be applied to yield a class of new radicals. With each property π of associative rings which satisfies the conditions of Theorem II. 7. 1 or II. 7. 2 we can associate a ***π.s.h.***-radical. We recall that a ring R is a ***π.s.h.***-ring if every homomorphic image of R does not contain a non zero π-ideal (i. e. is π-semi simple). In particular, Theorem II. 7. 2 can be applied to the following properties: quasi-regularity, nillity, semi-nilpotency and nilpotency.

In particular, a ring R is nilpotent-***s.h.*** if every homomorphic image does not contain nilpotent ideals. Now a regular ring is evidently a ring of this type so that nilpotent-***s.h.*** radical contains the maximal regular ideal defined in [5].

C. *The FI-radicals.* The notions of *I*-rings and *FI*-rings were introduced in [10] and the structure of such rings was determined there. An

associative ring R was called an I-ring if every non nil right ideal contains an idempotent, and R was called an FI-ring if every homomorphic image of R is an I-ring. Similarly we shall call a ring R an I_S-ring (an I_L-ring) if every right ideal which is not semi-nilpotent (which is not an L-ring) contains an idempotent. In the same way we define FI_S-rings and FI_L-rings.

We wish to show that the three properties: FI, FI_S and FI_L are SRZ-properties and satisfy the conditions II. (D_r). Furthermore, the last two properties satisfy also II. (F_r). The proof will be carried out only for the property FI_S since the proof for the other two properties is similar.

The property I_S satisfies II.(A) and a similar proof to that of [10] Lemma 5. 4 yields that I_S satisfies II. (C_1) and II. (C_2), hence by Lemma II. 7. 1 it follows that FI_S is an R-property. Evidently this property satisfies II. (Z). To prove the validity of II. (D_r) we consider a right ideal J in an FI_S-ring R. Let J/Q be a homomorphic image of J and let P/Q be a right ideal in J/Q. Now, PJ is a right ideal in R and RQJ is a two sided ideal. Since R is an FI_S-ring, R/RQJ is an I_S-ring, so that $(PJ, RQJ)/RQJ$ either contains an idempotent or it is semi-nilpotent. Put $M = PJ \cap RQJ$, then the last quotient is isomorphic with PJ/M. PJM is a two sided ideal in PJ, hence one readily verifies that PJ/PJM is also either semi-nilpotent or contains an idempotent. Since $JRQJ \subseteq JQJ \subseteq Q$ it follows that

$$PJM \subseteq PJ \cap PJRQJ \subseteq PJ \cap Q.$$

Furthermore,

$$(PJ \cap Q)^4 \subseteq PJ(RQJ \cap PJ) = PJM.$$

Hence $PJ/PJ \cap Q$ will also be semi-nilpotent or contain an idempotent respectively what PJ/PJM does. Consequently, $(PJ, Q)/Q$, and hence also P/Q, will have the same property, which proves (D_r). Evidently this implies II. (D), and by Theorem II. 5. 1 it follows that FI_S satisfies also II. (E). Consequently, FI_S is an SRZ-property.

We now prove the validity of (F_r). Let R be an FI_S-semi simple ring and let J be a right FI_S-ideal in R. Consider the ideal $J^* = (J, RJ)$. Let J^*/Q be a homomorphic image of J^* and let P/Q be a right ideal in J^*/Q. We shall prove that either P/Q contains an idempotent or it is semi-nilpotent. Evidently, the same will hold if we replace Q by J^*QJ^*; or, equivalently, we may assume that Q is an ideal in R. In this case one readily verifies that the semi-nilpotent radical N^*/Q of J^*/Q is an ideal in R/Q (by Corollary 2. 1). For every $j \varepsilon J$, $(jP, Q)/Q$ is a homomorphic image of jP. The latter is a right ideal in J, hence either it is semi-nilpotent or contains an idem-

potent. If the first case holds for every $j \in J$, then $(JP, Q)/Q$ must be semi nilpotent [9] and hence $JP \subseteq N^*$. Thus

$$P^2 \subseteq (J, RJ)P \subseteq (JP, RJP) \subseteq (N^*, RN^*) \subseteq N^*.$$

This proves that $(P^2, Q)/Q$ and hence also P/Q are semi-nilpotent. If this is not the case, $(jP, Q)/Q$ contains an idempotent e for some $j \in J$, i. e. $e \equiv jp(Q)$, $j \in J$ and $p \in P$. By replacing p by pe we may assume that $pe \equiv p(Q)$. Since $e^2 \not\equiv 0(Q)$, $pj \not\equiv 0(Q)$ and the latter belongs to P and satisfies $(pj)^2 = p(jp)j = pej = pj(Q)$. This proves that P/Q contains an idempotent. Thus J^* is an FI_S-ideal in R; hence the FI_S-semi simplicity of R yields $J^* = 0$ and therefore $J = 0$.

A radical similar to those discussed above is the following: call a ring R an *I_0-ring* if every non zero right ideal in R contains an idempotent, and call R an *FI_0-ring* if every homomorphic images of R are I_0-rings. As in the preceding case FI_0 can be shown to be an R-property, but this property does not satisfy the conditions II. (D_r) and II. (F_r). Yet we can prove that FI_0 satisfies II. (D) and the two conditions II. (G_1) and II. (G_2). Thus FI_0 and the preceding properties FI, FI_S and FI_L satisfy the main results on the radical, in particular Theorem II. 2. 2. Furthermore, FI_0 satisfies Theorem II. 4. 1 and the last two properties satisfy Theorem II. 6. 3.

To prove that FI_0 satisfies II. (D) we consider an FI_0-ring R and an ideal A in R. If A is not an FI_0-ring then A contains an ideal Q such that A/Q is not an I_0-ring and, evidently, A/AQA cannot be an I_0-ring. Hence we may assume that Q is an ideal in R. If J/Q is a right ideal in A/Q which does not contain an idempotent, then since R/Q is an I_0-ring, $JA \subseteq Q$. The set $\{x \mid x \in A, xA \subseteq Q\}$ is an ideal T in R. If $T \not\equiv 0 \bmod Q$ then it contains an idempotent $e \bmod Q$ but then $e^2 \in eA \subseteq Q$ which is impossible. This proves that $J \equiv 0 \bmod Q$; that is, the validity of II. (D). Before proceeding with the proof of the conditions II. (G_1) and II. (G_2) we show that the ring R_n is an I_0-ring if and only if R is such. Indeed if R_n is an I_0-ring, one readily verifies that R is also an I_0-ring. Let R be an I_0-ring and let J be a right ideal in R_n. Since R cannot have a left annihilator we have $JR_n \neq 0$. Hence, $J(c_{ii}R) \neq 0$ for some i. We can also see that the coefficients of c_{ii} of the matrices of $J(c_{ii}R)$ constitute a non zero right ideal in R. This ideal contains an idempotent e and one readily observes that the matrix of $Jc_{ii}R$ which is of the form $ec_{ii} + \cdots$ is an idempotent of the ideal J, q. e. d. We turn now to the proof of II. (G_1). Let R be an FI_0-ring.

[9] This proof holds only for the properties FI_S and FI_L and it is not known whether it is valid for FI.

Let Q be an ideal in R_n. Since $R_nQR_n = T_n$, where T is an ideal in R we have $R_n/R_nQR_n \cong (R/T)_n$. The latter is an I_0-ring since R/T is such, hence R_n/R_nQR_n and therefore, also R_n/Q is an I_0-ring which proves that R_n is an FI_0-ring. To prove II. (G_2) let R be an FI_0-semi simple ring. If R_n is not FI_0-semi simple then it contains an ideal Q which is an FI_0-ring. Now Q/R_nQR_n can be an I_0-ring only if $Q = R_nQR_n$. Let T be an ideal in R such that $R_nQR_n = T_n$; then for every ideal P in T we have $Q/P_n \cong (T/P)_n$. Since Q is an FI_0-ring, T/P must be an I_0-ring. This proves that T is an FI_0-ideal in R hence $T = 0$. Therefore $Q = 0$ which proves II. (G_2).

The properties FI (FI_S, FI_L and FI_0) are symmetric, in the sense that if a ring R is I-ring then also every non nil *left* ideal in R contains idempotency ([10]), and conversely.

We conclude with the remark that the following relation exists between these radicals: $FI(R) \supseteq FI_S(R) \supseteq FI_L(R) \supseteq FI_0(R)$ for every ring R, and all these radicals contain the maximal regular ideal of the ring R which was defined in [5]. The last follows from the fact that a regular ring is also an FI_0-ring.

HEBREW UNIVERSITY
AND THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] S. A. Amitsur, "A general theory for radicals, I. Radicals in complete lattices," *American Journal of Mathematics*, vol. 74 (1952), pp. 774-786.

[2] ———, "A general theory for radicals, II. Radicals in rings and bicategories," *ibid.*, vol. 76 (1954), pp. 100-125.

[3] R. Baer, "Radical ideals," *ibid.*, vol. 69 (1947), pp. 46-58.

[4] B. Brown and N. H. McCoy, "Some theorems on groups with applications to ring theory," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 302-311.

[5] ———, "The maximal regular ideal of a ring," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 165-171.

[6] N. Jacobson, "The radical and semi simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1950), pp. 300-320.

[7] J. Levitzki, "On the radical of a general ring," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 462-466.

[8] ———, "Prime ideals and the lower radical," *American Journal of Mathematics*, vol. 73 (1951), pp. 25-29.

[9] ———, "A note on prime ideals," *Riveon lemathematica* (Hebrew), vol. 5 (1951-2), pp. 1-3, see also *Mathematical Reviews*, vol. 13 (1), (1952), p. 8.

[10] ———, "On the structure of algebraic algebras and related rings," *Transactions of the American Mathematical Society*, vol. 74 (1953), p. 404.

[11] N. H. McCoy, "Prime ideals in general rings," *American Journal of Mathematics*, vol. 71 (1948), pp. 823-833.

PAIRS OF QUADRATIC EQUATIONS IN A FINITE FIELD.*

By L. CARLITZ.

1. Introduction. Let $Q(x) = Q(x_1, \cdots, x_r)$ denote a quadratic form with coefficients in the finite field $GF(q)$, q odd. If $\delta \neq 0$ is the discriminant of Q, then it is well-known that the number of solution of the equation $Q(\xi_1, \cdots, \xi_r) = \alpha$ $(\xi_i \in GF(q))$ is given by

$$q^{2s} + q^s \psi((-1)^s \delta \alpha) \qquad (r = 2s+1); \tag{1.1}$$

$$q^{2s-1} + q^{s-1} \psi((-1)^s \delta) k(\alpha) \qquad (r = 2s), \tag{1.2}$$

where $\psi(\alpha) = 0, +1, -1$ according as $\alpha = 0$, square or non-square of $GF(q)$, and $k(0) = q-1$, $k(\alpha) = -1$ for $\alpha \neq 0$.

If Q_1 and Q_2 denote two arbitrary quadratic forms then the question of the number of solutions of the system $Q_1(\xi_1, \cdots, \xi_r) = \alpha$, $Q_2(\xi_1, \cdots, \xi_r) = \beta$ seems to be much more difficult. In the present paper we obtain a number of results concerning such systems. In particular we consider the following problem. Let $Q_1, \cdots, Q_r$ denote r nonsingular quadratic forms with coefficients in $GF(q)$ in nonoverlapping sets of unknowns; let $\beta_1, \cdots, \beta_r$ be distinct and α, β arbitrary numbers of $GF(q)$. Let $N = N(\alpha, \beta)$ denote the number of solutions of the system

$$Q_1(\xi^{(1)}) + \cdots + Q_r(\xi^{(r)}) = \alpha, \qquad \beta_1 Q_1(\xi^{(1)}) + \cdots + \beta_r Q_r(\xi^{(r)}) = \beta, \tag{1.3}$$

where $\xi^{(i)}$ stands for $(\xi_1^{(i)}, \cdots, \xi_{s_i}^{(i)})$. Thus the total number of unknowns is

$$s = s_1 + \cdots + s_r. \tag{1.4}$$

The case $r = 1$ is trivial. Also the case $r = 2$ may be reduced to

$$(\beta_2 - \beta_1) Q_1(\xi_1^{(1)}) = \alpha\beta_2 - \beta, \qquad (\beta_1 - \beta_2) Q_2(\xi^{(2)}) = \alpha\beta_1 - \beta,$$

so that the number of solutions of the system may be found by means of (1.1) and (1.2). We may accordingly suppose $r \geqq 3$.

Fairly simple results for $N(\alpha, \beta)$ are obtained if certain additional hypotheses are introduced. For example if all the s_i are even then

$$\begin{aligned} Q^2 N(\alpha, \beta) &= q^s + \tau q^{\frac{1}{2}s} \psi((-1)^s) \{ \Big(\sum_{i=1}^{r} \psi((-1)^{\frac{1}{2}s_i} \delta_i) q^{\frac{1}{2}s_i} - 1 \Big) k(\beta - \alpha\beta_i) \\ &\qquad + k(\alpha) k(\beta) + k(\alpha) + k(\beta) \}, \end{aligned} \tag{1.5}$$

* Received April 13, 1953.

where δ_i is the discriminant of Q_i and $\tau = \prod \psi(\delta_i)$, while $k(\alpha)$ has the same meaning as in (1.2). For the general case see Theorems 3-6 below. We find that in certain cases the number of solutions is expressed in terms of the sum

$$(1.6) \qquad \sum_\lambda \psi((\lambda-\beta_1)\cdots(\lambda-\beta_t)), \qquad \lambda \in GF(q),$$

where it is assumed that $s_1, \cdots, s_t$ are odd while the remaining s_i are even. Now in some cases the sum (1.6) can be evaluated explicitly. For example Jacobsthal [5] has discussed the sum $\sum \psi(\lambda(\lambda^2-1))$ in the case $q=p$. It is easy to extend his result to $GF(q)$ and thus get simple explicit formulas for $N(\alpha,\beta)$ in certain cases (see Theorem 7 and 8 below). It may be remarked that (1.6) can be evaluated in some additional cases. For a systematic discussion of Jacobsthal sums see Whiteman [8], [9]; we shall however not apply these results in the present paper.

In the next place we discuss a conjecture of André Weil [7] for the homogeneous system (1.3) with $\alpha=\beta=0$. Let $N^{(m)}$ denote the number of solutions in the enlarged field $GF(q^m)$; let N_m denote the number of "rational points," where $(q-1)N_m = N^{(m)}-1$, then we verify that $\sum N_m u^{m-1} = d\log Z(u)/du$, $(m \geqq 1)$, where $Z(u)$ is a certain rational function.

Finally we discuss briefly the case $q=2^n$. We now suppose

$$(1.7) \qquad Q(u) = \sum_{i\leqq j} \alpha_{ij}u_iu_j \qquad (\alpha_{ij} \in GF(2^n))$$

and use some results of Dickson [3], § 199, on normal forms for (1.7). In particular if $Q(u)$ is of even rank then it possesses a certain invariant δ. We find that if each Q_i is of even rank s_i then the number of solutions of (1.3) satisfies

$$(1.8) \quad q^2N(\alpha,\beta) = q^s + lq^{\frac{1}{2}s}\{\sum_{i=1}^{r}(q^{\frac{1}{2}s_i}l(\delta_i)-1)k(\beta-\alpha\beta_i) + k(\alpha)k(\beta)+k(\alpha)+k(\beta)\},$$

where $l(0)=1$, $l(\delta)=-1$ for $\delta\neq 0$, $l=\prod l(\delta_i)$; note that (1.8) and (1.5) are very similar. If some of the s_i are odd the final results are somewhat simpler (see (9.19), (9.20)).

2. Preliminaries. If $q=p^n$, put

$$(2.1) \qquad e(\alpha) = e^{2\pi i t(\alpha)/p}, \qquad t(\alpha) = \alpha + \alpha^p + \cdots + \alpha^{p^{n-1}}.$$

Then it is familiar that

$$(2.2) \qquad \sum_\beta e(\alpha\beta) = q \quad (\alpha=0), \qquad \sum_\beta e(\alpha\beta) = 0 \quad (\alpha\neq 0),$$

where the summation is over all $\beta \in GF(q)$. It follows from (2.2) that the function $k(\alpha)$ occurring in (1.2) satisfies

$$k(\alpha) = \sum_{\beta \neq 0} e(\alpha\beta). \tag{2.3}$$

We next define the Gauss sum [1], § 3,

$$G(\alpha) = \sum_{\beta} e(\alpha\beta^2) \quad (\alpha \neq 0); \text{ so that } G(\alpha) = \psi(\alpha)G(1), \tag{2.4, 2.5}$$

where $\psi(\alpha)$ has the same meaning as in (1.1) and (1.2); also

$$G^2(1) = \psi(-1)q; \tag{2.6}$$

$$\sum_{\beta} \psi(\beta) e(\alpha\beta) = G(\alpha), \tag{2.7}$$

valid for all α (including $\alpha = 0$). If $Q(\xi_1, \cdots, \xi_s)$ is a quadratic form we put

$$G(Q) = \sum e(Q(\xi_1, \cdots, \xi_r)); \text{ hence } G(Q) = \psi(\delta)G_1^r \quad (\delta \neq 0), \tag{2.8, 2.9}$$

if δ is the discriminant of Q. Making use of (2.2) it is evident that the number of solutions of (1.3) satisfies

$$q^2N = \sum_{\lambda,\mu} \sum_{\xi} e\{\lambda \sum_{i=1}^{r} Q_i(\xi^{(i)}) - \lambda\alpha\} e\{\mu \sum_{i=1}^{r} \beta_i Q_i(\xi^{(i)}) - \mu\beta\}, \tag{2.10}$$

where λ, μ independently range over the numbers of $GF(q)$ and $\sum_{\xi}$ indicates a summation over the s numbers $\xi_j^{(i)}$ $(j = 1, \cdots, s_i)$. Using (2.9) we may rewrite the right member of (2.10) as

$$\sum_{\lambda,\mu} e(-\lambda\alpha + \mu\beta) \prod_{i=1}^{r} G\{(\lambda - \mu\beta_i)Q_i\}. \tag{2.11}$$

In order to simplify this expression we break the sum into several parts. In the first place the term $\lambda = \mu = 0$ contributes q^s. Next if $\lambda = \mu\beta_i$ for some fixed i and $\mu \neq 0$ we get

$$q^{s_i} \sum_{\mu \neq 0} e((\beta - \alpha\beta_i)\mu) \prod_{j=1}^{r}{}' G\{(\beta_i - \beta_j)\mu Q_j\}$$

$$= q^{s_i} G^{s-s_i}(1) \sum_{\mu \neq 0} \psi(\mu^{s-s_i}) e((\beta - \alpha\beta_i)\mu) \prod_{j=1}^{r}{}' \psi((\beta_i - \beta_j)^{s_j}\delta_j),$$

where δ_j is the discriminant of Q_j and the prime refers to the exclusion of $j = i$. Thus the total contribution of these terms is

$$\sum_{i=1}^{r} \tau_i q^{s_i} G^{s-s_i}(1) \sum_{\mu \neq 0} \psi(\mu^{s-s_i}) e((\beta - \alpha\beta_i)\mu), \tag{2.12}$$

$$\tau_i = \prod_{j=1}^{r}{}' \psi((\beta_i - \beta_j)^{s_j}\delta_j), \tag{2.13}$$

where $j \neq i$. The remaining terms in (2.11) give

$$G^s(1) \sum_{\lambda,\mu}{}' e(-\lambda\alpha + \mu\beta) \prod_{i=1}^{r} \psi((\lambda - \mu\beta_i)^{s_i}\delta_i), \text{ where } \lambda \neq \mu\beta_i. \tag{2.14}$$

Substituting from (2.12) and (2.14) it is clear that (2.10) becomes

$$\begin{aligned} q^2N = q^s + \sum_{i=1}^{r} \tau_i q^{s_i} G^{s-s_i}(1) \sum_{\mu} \psi(\mu^{s-s_i})\, e((\beta - \alpha\beta_i)\mu) \\ + G^s(1) \sum_{\lambda,\mu} e(-\lambda\alpha + \mu\beta) \prod_{i=1}^{r} \psi((\lambda - \mu\beta_i)^{s_i}\delta_i); \end{aligned} \tag{2.15}$$

the summation may now be extended over all λ, μ since the additional terms vanish. Further simplification of (2.15) depends on the introduction of certain additional hypotheses concerning α, β or the Q_j.

3. The case s_i even. If all the s_i are even then, first of all, (2.13) becomes

$$\tau_i = \prod_{j=1}^{r}{}' \psi(\delta_j) = \psi(\delta_i)\tau, \ (j \neq i), \qquad \tau = \prod_{j=1}^{r} \psi(\delta_j). \tag{3.1}$$

As for (2.15) it evidently reduces to

$$\begin{aligned} q^2N = q^s + \sum_{i=1}^{r} \tau_i q^{s_i} G^{s-s_i}(1) \sum_{\mu \neq 0} e((\beta - \alpha\beta_i)\mu) \\ + \tau G^s(1) \sum_{\lambda \neq \mu\beta_i} e(-\lambda\alpha + \mu\beta), \end{aligned} \tag{3.2}$$

where τ, τ_i are defined by (3.1) and the last sum is over all λ, μ such that $\lambda \neq \mu\beta_i$ for any i. This sum evidently is

$$\begin{aligned} \sum_{\lambda,\mu} e(-\lambda\alpha + \mu\beta) - \sum_{i=1}^{r} \sum_{\mu \neq 0} e((\beta - \alpha\beta_i)\mu) - 1 \\ = (k(\alpha) + 1)(k(\beta) + 1) - \sum_{i=1}^{r} k(\beta - \alpha\beta_i) - 1, \end{aligned}$$

where we have used (2.3). Thus (3.2) becomes

$$\begin{aligned} q^2N = q^s + \tau \sum_{i=1}^{r} \{\psi(\delta_i) q^{s_i} G^{s-s_i}(1) - G^s(1)\} k(\beta - \alpha\beta_i) \\ + \tau G^s(1)(k(\alpha)k(\beta) + k(\alpha) + k(\beta)). \end{aligned} \tag{3.3}$$

Finally by means of (2.6) we can eliminate $G(1)$ and (3.3) takes on the following explicit form.

$$\begin{aligned} q^2N(\alpha,\beta) = q^s + \tau\psi((-1)^{\frac{1}{2}s})q^{\frac{1}{2}s} \sum_{i=1}^{r} \{\psi((-1)^{\frac{1}{2}s_i}\delta_i)q^{\frac{1}{2}s_i} - 1\} k(\beta - \alpha\beta_i) \\ + \tau\psi((-1)^{\frac{1}{2}s})q^{\frac{1}{2}s}(k(\alpha)k(\beta) + k(\alpha) + k(\beta)). \end{aligned} \tag{3.4}$$

We may now state

THEOREM 1. *If all the s_i are even then the number of solutions of the system* (1.3) *is given by* (3.4), *where δ_i is the discriminant of Q_i and* $\tau = \prod_i \psi(\delta_i)$.

This result may be compared with (1.2). The special case $\alpha = \beta = 0$ is of some intereste. We state

THEOREM 2. *If all the s_i are even and in addition $\alpha = \beta = 0$ then the number of solutions of* (1.3) *satisfies*

$$q^2N(0,0) = q^s + \tau\psi((-1)^{\frac{1}{2}s})q^{\frac{1}{2}s}(q-1) \times \{\sum_{i=1}^{r} \psi((-1)^{\frac{1}{2}s_i}\delta_i)q^{\frac{1}{2}s_i} + q + 1 - r\}. \tag{3.5}$$

The following special case of Theorem 1 may be mentioned. The number of solutions of the system $\xi_1\eta_1 + \cdots + \xi_r\eta_r = \alpha$, $\beta_1\xi_1\eta_1 + \cdots + \beta_r\xi_r\eta_r = \beta$ is equal to

$$q^{2r-2} + q^{r-2}(q-1)\sum_{i=1}^{r} k(\beta - \alpha\beta_i) + q^{r-2}(k(\alpha)k(\beta) + k(\alpha) + k(\beta)).$$

As a partial check of (3.4) and (3.5) we remark that for $r = 2$ the two formulas are in agreement with (1.2).

It may also be worth noting that for $\alpha = 0$, $\beta \neq 0$, (3.4) becomes

$$q^2N(0,\beta) = q^s + \tau\psi((-1)^{\frac{1}{2}s})q^{\frac{1}{2}s}\sum_{i=1}^{r} \psi((-1)^{\frac{1}{2}s_i}\delta_i)q^{\frac{1}{2}s_i} \qquad (\beta \neq 0). \tag{3.6}$$

4. The general case. If one or more of the s_i is odd, the situation is rather different. The final results depend on whether the number of odd s_i is itself odd or even. Let us assume the s_i so numbered that $s_1, \cdots, s_t$ are odd, $1 \leqq t \leqq r$, while $s_{t+1}, \cdots, s_r$ are even. In the double sum in (2.15) the terms for which $\mu = 0$ contribute

$$\sum_{\lambda} e(-\lambda\alpha)\psi(\lambda^t)\tau \qquad (\tau = \prod_{i=1}^{r} \psi(\delta_i)). \tag{4.1}$$

To evaluate this sum we use (2.3) or (2.7) according as t is even or odd. It follows that (4.1) is equal to

$$-k(\alpha)\tau \ (t \text{ even}) \text{ or } \psi(-\alpha)G(1)\tau \ (t \text{ odd}). \tag{4.2}$$

The remaining terms in the double sum in (2.15) give (on replacing λ by $\lambda\mu$)

$$\tau\sum_{\lambda}\sum_{\mu\neq 0} e((-\lambda\alpha + \beta)\mu)\psi(\mu^t)\prod_{i=1}^{t}\psi(\lambda - \beta_i)\prod_{j=t+1}^{r}\psi^2(\lambda - \beta_j). \tag{4.3}$$

Again using (2.3) and (2.7) we find that (4.3) reduces to

$$\tau \sum_{\lambda} k(-\lambda\alpha+\beta)T(\lambda) \quad (t \text{ even}) \text{ or } \tau G(1) \sum_{\lambda} \psi(-\lambda\alpha+\beta)T(\lambda) \quad (t \text{ odd}), \tag{4.4}$$

where

$$T(\lambda) = \prod_{i=1}^{t} \psi(\lambda-\beta_i) \prod_{j=t+1}^{r} \psi(\lambda-\beta_j)^2. \tag{4.5}$$

Turning next to the sum

$$\sum_{\mu} \psi(\mu^{s-s_i}) e((\beta-\alpha\beta_i)\mu) \tag{4.6}$$

occurring in (2.15), we remark that $s \equiv t \pmod 2$, so that (4.6) reduces to

$$k(\beta-\alpha\beta_i) \quad (i \leqq t \text{ for } t \text{ odd}, i > t \text{ for } t \text{ even}) \text{ or } G(\beta-\alpha\beta_i) \text{ (otherwise)}. \tag{4.7}$$

Let K_t denote the sum $\sum_{i}$ in (2.15); it follows from (4.7) that

$$\begin{aligned} K_t &= \sum_{i=1}^{t} \tau_i q^{s_i} G^{s-s_i}(1) k(\beta-\alpha\beta_i) + \sum_{i=t+1}^{r} \tau_i q^{s_i} G^{s-s_i}(1) G(\beta-\alpha\beta_i) \quad (t \text{ odd}) \\ K_t &= \sum_{i=1}^{t} \tau_i q^{s_i} G^{s-s_i}(1) G(\beta-\alpha\beta_i) + \sum_{i=t+1}^{r} \tau_i q^{s_i} G^{s-s_i}(1) k(\beta-\alpha\beta_i) \quad (t \text{ even}), \end{aligned} \tag{4.8}$$

where by (2.13)

$$\tau_i = \tau\psi(\delta_i) \prod_{j=1}^{t}{}' \psi(\beta_i-\beta_j) \quad (i \leqq t) \text{ or } \tau_i = \tau\psi(\delta_i) \prod_{j=1}^{t} \psi(\beta_i-\beta_j) \quad (i > t), \tag{4.9}$$

the prime indicates $j \neq i$ and as before

$$\tau = \prod_{i=1}^{r} \psi(\delta_i). \tag{4.10}$$

If we now collect the terms corresponding to (4.2), (4.4), (4.8) we see that (2.15) reduces to

$$\begin{aligned} q^2N &= q^s + K_t + \{k(\alpha) + \sum_{\lambda} k(-\lambda\alpha+\beta)T(\lambda)\}\tau G^s(1) \quad (t \text{ even}) \\ q^2N &= q^s + K_t + \{\psi(-\alpha) + \sum_{\lambda} \psi(-\lambda\alpha+\beta)T(\lambda)\}\tau G^{s+1}(1) \quad (t \text{ odd}). \end{aligned} \tag{4.11}$$

This proves

THEOREM 3. *If $s_1, \cdots, s_t$ are odd while $s_{t+1}, \cdots, s_r$ are even then the number of solution of (1.3) satisfies (4.11), where K_t, $T(\lambda)$ are defined by (4.8) and (4.5), respectively, and τ is defined by (4.10).*

If $t = 0$, (4.11) should coincide with (3.4). Since by (4.5) and (4.8)

$$K_0 = \tau \sum_{i=1}^{r} \psi(\delta_i) q^{s_i} G^{s-s_i}(1) k(\beta-\alpha\beta_i)$$

and $k(\alpha)+\sum_\lambda k(-\lambda\alpha+\beta)$ is $k(\alpha)$ if $\alpha\neq 0$ and $k(\alpha)+k(\beta)q$ if $\alpha=0$, it follows that the first half of (4.11) does indeed reduce to (3.4); this furnishes a partial check for (4.11).

In the second place if $t=r=2$,

$$K_2=\psi((\beta_1-\beta_2)\delta_2)q^{s_1}G^{s_2}(1)G(\beta-\alpha\beta_1)+\psi((\beta_2-\beta_1)\delta_1)q^{s_2}G^{s_1}(1)G(\beta-\alpha\beta_2).$$

Also
$$\begin{aligned}k(\alpha)+\sum_\lambda k(-\lambda\alpha+\beta)&\psi((\lambda-\beta_1)(\lambda-\beta_2))\\&=-1-\sum_\lambda\psi((\lambda-\beta_1)(\lambda-\beta_2))+q\psi((\beta-\alpha\beta_1)(\beta-\alpha\beta_2))\\&=q\psi((\beta-\alpha\beta_1)(\beta-\alpha\beta_2))\end{aligned}$$

for $\alpha\neq 0$, while for $\alpha=0$ we get $q-1-k(\beta)=q\psi(\beta^2)$. Thus the right member of (4.11) becomes

$$\{q^{s_1}+q^{\frac{1}{2}(s+1)}\psi((-1)^{\frac{1}{2}(s_1-1)}\delta_1(\alpha\beta_2-\beta)(\beta_2-\beta_1)\}\cdot\{q^{s_2}+q^{\frac{1}{2}(s_2+1)}\psi((-1)^{\frac{1}{2}(s_2-1)}\delta_2(\alpha\beta_1-\beta)(\beta_1-\beta_2)\},$$

in agreement with (1.2).

5. Some special cases. If $\alpha\neq 0$ we may replace the second equation in (1.3) by $(\beta-\alpha\beta_1)Q_1(\xi^{(1)})+\cdots+(\beta-\alpha\beta_r)Q_r(\xi^{(r)})=0$; in other words there is no loss in generality in assuming that $\beta=0$. With this normalization, (4.11) becomes

$$q^2N(\alpha,0)=q^s+K_t(\alpha,0)+\{k(\alpha)+\sum_\lambda k(-\lambda\alpha)T(\lambda)\}\tau G^s(1) \quad (t \text{ even}),$$

$$q^2N(\alpha,0)=q^s+K_t(\alpha,0)+\{\psi(-\alpha)+\sum_\lambda\psi(-\lambda\alpha)T(\lambda)\}\tau G^{s+1}(1) \quad (t \text{ odd}), \tag{5.1}$$

where $T(\lambda)$ is defined by (4.5),

$$K_t(\alpha,0)=\sum_{i=1}^{t}\tau_iq^{s_i}G^{s-s_i}(1)k(-\alpha\beta_i)+\sum_{i=t+1}^{r}\tau_iq^{s_i}G^{s-s_i}(1)G(-\alpha\beta_i) \quad (t \text{ odd}),$$

$$K_t(\alpha,0)=\sum_{i=1}^{t}\tau_iq^{s_i}G^{s-s_i}(1)G(-\alpha\beta_i)+\sum_{i=t+1}^{r}\tau_iq^{s_i}G^{s-s_i}(1)k(-\alpha\beta_i) \quad (t \text{ even}), \tag{5.2}$$

and τ_i is defined by (4.9). The formulas (5.1), (5.2) are valid for all α. This proves

THEOREM 4. *If $s_1,\cdots,s_t$ are odd while $s_{t+1},\cdots,s_r$ are even and $\beta=0$ then the number of solutions $N(\alpha,0)$ of* (1.3) *is determined by* (5.1) *and* (5.2) *together with* (4.5) *and* (4.9).

As already pointed out if $\alpha \neq 0$ then Theorem 4 involves no specialization. However, the case $\alpha = 0$ is not covered (see Theorem 6 below).

In particular if $\alpha = 0$ it is easily verified that (5. 1) and (5. 2) reduce to

$$q^2N(0,0) = q^s + K_t(0,0) + (q-1)\{1 + \sum_{\lambda} T(\lambda)\}\tau G^s(1) \tag{5.3}$$

$$\text{or } q^2N(0,0) = q^s + K_t(0,0)$$

according as t is even or odd, where, in the respective cases,

$$K_t(0,0) = (q-1)\sum_{i=t+1}^{r} \tau_i q^{s_i} G^{s-s_i}(1) \tag{5.4}$$

$$\text{or } K_t(0,0) = (q-1)\sum_{i=1}^{t} \tau_i q^{s_i} G^{s-s_i}(1)$$

(τ_i is defined by (4. 9)). This proves

THEOREM 5. *If $s_1, \cdots, s_t$ are odd while $s_{t+1}, \cdots, s_r$ are even and $\alpha = \beta = 0$, then the number of solutions $N(0,0)$ of the system* (1. 3) *is determined by* (5. 3) *and* (5. 4).

Returning to (4. 11) let us take $\alpha = 0$; since the case $\alpha = \beta = 0$ is covered by the last theorem we may assume $\beta \neq 0$. We find that (4. 11) reduces to

$$\begin{aligned} q^2N(0,\beta) &= q^s + K_t(0,\beta) + \{q - 1 - \sum_{\lambda} T(\lambda)\}\tau G^s(1) \quad (t \text{ even}) \\ q^2N(0,\beta) &= q^s + K_t(0,\beta) + \sum_{\lambda} T(\lambda)\psi(\beta)\tau G^{s+1}(1) \quad (t \text{ odd}), \end{aligned} \tag{5.5}$$

where

$$\begin{aligned} K_t(0,\beta) &= \sum_{i=1}^{t} \tau_i q^{s_i} G^{s-s_i}(1) + \sum_{i=t+1}^{r} \tau_i q^{s_i} G^{s-s_i}(1) G(\beta) \quad (t \text{ odd}) \\ K_t(0,\beta) &= \sum_{i=1}^{t} \tau_i q^{s_i} G^{s-s_i}(1) G(\beta) - \sum_{i=t+1}^{r} \tau_i q^{s_i} G^{s-s_i}(1) \quad (t \text{ even}) \end{aligned} \tag{5.6}$$

and τ_i and τ are determined by (4. 9) and (4. 10). We state

THEOREM 6. *If $\alpha = 0$, $\beta \neq 0$, the number of solutions $N(0,\beta)$ of the system* (1. 3) *is determined by* (5. 5) *and* (5. 6).

When $t = r$ the results of this section simplify and it may be worth while stating the specialized results explicitly. In the first place we now have

$$T(\lambda) = \prod_{i=1}^{t} (\lambda - \beta_i), \qquad \tau_i = \tau\psi(\delta_i) \prod_{j=1}^{t}{}' (\beta_i - \beta_j) \qquad (j \neq i). \tag{5.7}$$

In the next place, according as t is even or odd,

$$(5.8)\quad K_t(\alpha,0) \text{ is } \sum_{i=1}^{t}\tau_i q^{s_i}G^{s-s_i}(1)G(-\alpha\beta_i) \text{ or } \sum_{i=1}^{t}\tau_i q^{s_i}G^{s-s_i}(1)k(-\alpha\beta_i);$$

then $N(\alpha,0)$ is determined by (5.1) together with (5.7) and (5.8).

If $\alpha=0$, $\beta\neq 0$, then, according as t is even or odd,

$$(5.9)\qquad K_t(0,\beta) \text{ is } \sum_{i=1}^{t}\tau_i q^{s_i}G^{s-s_i}(1)G(\beta) \text{ or } -\sum_{i=1}^{t}\tau_i q^{s_i}G^{s-s_i}(1)$$

and $N(0,\beta)$ is determined by (5.7) and (5.9).

6. The sum $\sum T(\lambda)$. Further simplification of (5.1) or (5.5) depends upon the evaluation of the sums

$$(6.1)\qquad \sum_t k(-\lambda\alpha)T(\lambda), \qquad \sum_\lambda \psi(-\lambda\alpha)T(\lambda), \qquad \sum_\lambda T(\lambda),$$

where $T(\lambda)$ is defined by (4.5). We begin with the last of the sums (6.1). Put

$$(6.2)\qquad J_t=\sum_\lambda T_1(\lambda), \qquad T_1(\lambda)=\prod_{i=1}^{t}\psi(\lambda-\beta_i);$$

then

$$(6.3)\qquad J_t=\sum_\lambda T(\lambda)+\sum_{i=t+1}^{r}\prod_{j=1}^{t}\psi(\beta_i-\beta_j).$$

Now it is clear to begin with that

$$(6.4)\qquad J_1=0, \qquad\qquad (6.5)\qquad J_2=-1.$$

In the case $t=3$ the general sum can be reduced to one of the type

$$(6.6)\qquad f(\beta)=\sum_\lambda \psi(\lambda(\lambda-1)(\lambda-\beta))$$

with $\beta\neq 0$ or 1. Consider

$$\begin{aligned}\sum_\beta f^2(\beta) &= \sum_\lambda \psi(\lambda(\lambda-1))\sum_\mu \psi(\mu(\mu-1))\sum_\beta \psi((\lambda-\beta)(\mu-\beta))\\ &= \sum_\lambda \psi^2(\lambda(\lambda-1))\sum_\beta \psi^2(\lambda-\beta)\\ &\qquad +\sum_{\lambda\neq\mu}\psi(\lambda(\lambda-1))\psi(\mu(\mu-1))\sum_\beta \psi((\lambda-\beta)(\mu-\beta))\\ &= (q-2)q-\sum_{\lambda,\mu}\psi(\lambda(\lambda-1))\psi(\mu(\mu-1))=q^2-2q-1,\end{aligned}$$

by (6.5). Since $f(0)=\sum_\lambda \psi(\lambda^2(\lambda-1))=\sum_\lambda \psi(\lambda-1)-\psi(-1)=-\psi(-1)$ and similarly $f(1)=-\psi(-1)$, it follows that

$$(6.7)\qquad \sum_{\beta\neq 0,1} f^2(\beta)=q^2-2q-3.$$

In the next place, generalizing [5] slightly, we put

$$\phi(\beta) = \sum_{\lambda} \psi(\lambda(\lambda^2 + \beta)) \qquad (\beta \neq 0); \tag{6.8}$$

then

$$\phi(\alpha^2\beta) = \psi(\alpha)\phi(\beta) \qquad (\alpha \neq 0), \tag{6.9}$$

and therefore

$$\sum_{\beta \neq 0} \phi^2(\beta) = \tfrac{1}{2}(q-1)(\phi^2(1) + \phi^2(\gamma)), \tag{6.10}$$

where γ is any number of $GF(q)$ such that $\psi(\gamma) = -1$. On the other hand exactly as in the proof of (6.7) we have

$$\sum_{\beta \neq 0} \phi^2(\beta) = (1 + \psi(-1))(q-1)q. \tag{6.11}$$

Comparison with (6.10) leads at once to

$$\phi^2(1) + \phi^2(\psi) = 2(1 + \psi(-1))q. \tag{6.12}$$

It is evident from (6.12) that

$$\phi(\beta) = 0 \qquad (q \equiv 3 \pmod 4); \tag{6.13}$$

$$(\tfrac{1}{2}\phi(1))^2 + (\tfrac{1}{2}\phi(\gamma))^2 = q \qquad (q \equiv 1 \pmod 4). \tag{6.14}$$

It is clear from (6.8) that $\phi(\beta)$ is even for all β; also since

$$\sum_{\lambda} (1 - \psi(\lambda))(1 - \psi(\lambda + 1))(1 - \psi(\lambda - 1)) = q - 3 - \phi(-1) \equiv 0 \pmod 8,$$

it follows from (6.9) that

$$\tfrac{1}{2}\phi(1) \equiv -1 \pmod 4. \tag{6.15}$$

We may restate (6.14) and (6.15) in the following way. Put $a = \frac{1}{2}\phi(1)$, $b = \frac{1}{2}\phi(\gamma)$, then

$$a^2 + b^2 = q, \qquad a \equiv -1 \pmod 4. \tag{6.16}$$

Moreover (6.8) implies

$$\phi(\beta) = \sum_{\lambda} \sum_{i=0}^{\frac{1}{2}(q-1)} \binom{\frac{1}{2}(q-1)}{i} \beta^i \lambda^{\frac{1}{2}(q-1)-2i} = -\binom{\frac{1}{2}(q-1)}{\frac{1}{4}(q-1)} \beta^{\frac{1}{4}(q-1)}$$

and therefore using a known property of binomial coefficients

$$\phi(\beta) \equiv -\binom{\frac{1}{2}(p-1)}{\frac{1}{4}(p-1)}^n \psi_4(\beta) \pmod p \qquad (q = p^n); \tag{6.17}$$

$$2a = \phi(1) \equiv -\binom{\frac{1}{2}(p-1)}{\frac{1}{4}(p-1)}^n \pmod p, \tag{6.18}$$

where $\psi_4(-\beta) = \beta^{\frac{1}{4}(q-1)}$. It is clear that a is uniquely determined by means of (6.16) and the first of (6.18). We have thus generalized Jacobsthal's theorem.

In the next place we have the easily verified relation [4], p. 227,

$$\sum_{\lambda} \psi((\lambda-\beta_1)\cdots(\lambda-\beta_4)) = \psi(\eta)\sum_{\lambda}\psi(\lambda(\lambda+1)(\lambda+\rho)), \tag{6.19}$$

where the β's are distinct, $\eta = (\beta_1-\beta_3)/(\beta_1-\beta_4)$ and

$$\rho = (\beta_1\beta_2, \beta_3\beta_4) = (\beta_1-\beta_3)/(\beta_1-\beta_4) : (\beta_2-\beta_3)/(\beta_2-\beta_4), \tag{6.20}$$

the cross-ratio of $\beta_1, \beta_2, \beta_3, \beta_4$. In particular if $\rho = -1$, then (6.19) becomes

$$\begin{aligned}\sum_{\lambda} \psi((\lambda-\beta_1)\cdots(\lambda-\beta_4)) &= \psi(\eta)\sum_{\lambda}\psi(\lambda(\lambda^2-1)) \\ &= (-1)^{\frac{1}{4}(q-1)}\psi(\eta)\phi(1),\end{aligned} \tag{6.21}$$

by (6.8) and (6.9). Then when $\beta_1, \beta_2, \beta_3, \beta_4$ form a harmonic range the sum J_4 (defined in (6.2)) is expressed in terms of $\phi(1)$.

Returning to (6.1) it is apparent that each of the first two sums can be expressed in terms of the third. Thus

$$\sum_{\lambda} k(-\lambda\alpha)T(\lambda) = -\sum_{\lambda} T(\lambda) + qT(0) \qquad (\alpha \neq 0) \tag{6.22}$$

and

$$\sum_{\lambda} \psi(-\lambda\alpha) = \psi(-\alpha)\sum_{\lambda} T^*(\lambda), \qquad T^*(\lambda) = \psi(\lambda)T(\lambda). \tag{6.23}$$

We also remark that by a theorem of Weil [6]

$$\sum_{\lambda} T(\lambda) = O(q^{\frac{1}{2}}) \qquad (q \to \infty, t \geqq 1). \tag{6.24}$$

For a systematic treatment of Jacobsthal sums, see [8], [9].

7. Applications. In view of the results of § 6, some of the formulas of § 5 can be made more explicit. For example if $t = 4$, (6.3) becomes

$$\sum_{\lambda} T(\lambda) = J_4 + \sum_{i=5}^{r}\prod_{j=1}^{4}\psi(\beta_i-\beta_j), \tag{7.1}$$

where

$$J_4 = \sum_{\lambda}\psi((\lambda-\beta_1)\cdots(\lambda-\beta_4)). \tag{7.2}$$

We assume that the cross-ratio

$$(\beta_1\beta_2, \beta_3\beta_4) = -1. \tag{7.3}$$

It follows from (6.21) that

$$J_4 = 2a(-1)^{\frac{1}{4}(q-1)}\psi(\eta) \qquad (q \equiv 1 \pmod 4), \tag{7.4}$$

where a is determined by means of (6.16), while

$$J_4 = 0 \qquad (q \equiv 3 \pmod 4). \tag{7.5}$$

On the other hand if $t = 3$ and

$$2\beta_1 = \beta_2 + \beta_3, \tag{7.6}$$

then we have

$$J_3 = \sum_\lambda \psi((\lambda - \beta_1)(\lambda - \beta_2)(\lambda - \beta_3)) = \psi(\beta_1 - \beta_2) \sum_\lambda \psi(\lambda(\lambda^2 - 1)),$$

so that

$$J_3 = 2a(-1)^{\frac{1}{4}(q-1)}\psi(\beta_1 - \beta_2) \qquad (q \equiv 1 \pmod 4); \tag{7.7}$$

$$J_3 = 0 \qquad (q \equiv 3 \pmod 4).$$

Let us now return to (5.1). We have

$$\sum_\lambda k(-\lambda\alpha)T(\lambda) = k(\alpha)\sum_\lambda T(\lambda) + (q - 1 - k(\alpha))T(0). \tag{7.8}$$

Then if $t = 4$ and (7.3) is satisfied, the sum in the right member of (7.8) is evaluated by means of (7.1) and (7.4) or (7.5). If $t = 3$ we may assume $\alpha \neq 0$; then

$$\sum_\lambda \psi(-\lambda\alpha)T(\lambda) = \psi(-\alpha)\sum_\lambda \psi(\lambda)T(\lambda).$$

Since $\psi(\lambda)T_1(\lambda) = \psi(\lambda(\lambda - \beta_1)(\lambda - \beta_2)(\lambda - \beta_3))$, the condition (7.3) reduces to

$$(\beta_1 - \beta_3)/(\beta_2 - \beta_3) = -\beta_1/\beta_2 \tag{7.9}$$

(provided $\beta_1\beta_2\beta_3 \neq 0$). Then the sum $\sum \psi(\lambda)T(\lambda)$ is determined by means of (7.4) or (7.5). If $\beta_3 = 0$ we get

$$\sum \psi(\lambda)T_1(\lambda) = \sum_\lambda \psi(\lambda^2(\lambda - \beta_1)(\lambda - \beta_2)) = -1 - \psi(\beta_1\beta_2).$$

If $t = r$ the above results are particularly compact. We state

THEOREM 7. *If $r = t = 4$ and (7.3) holds then $N(\alpha, 0)$ satisfies*

$$q^2N(\alpha, 0) = q^s + K_4(\alpha, 0) + \{k(\alpha)(J_4 + 1) + q - 1 - k(\alpha)\psi(\beta_1\beta_2\beta_3\beta_4)\}\tau G^s(1), \tag{7.10}$$

where J_4 is determined by (7.4) or (7.5).

If $r = t = 3$ and (7.9) holds then $N(\alpha, 0)$ satisfies

$$q^2 N(\alpha, 0) = q^s + K_3(\alpha, 0) + \psi(-\alpha)(J_4 + 1)\tau G^{s+1}(1), \tag{7.11}$$

but if $\beta_1\beta_2\beta_3 = 0$ then

$$q^2 N(\alpha, 0) = q^s + K_3(\alpha, 0) - \psi(-\alpha\beta_1\beta_2)\tau G^{s+1}(1). \tag{7.12}$$

In all cases K_4, K_3 are determined by (5.2).

Turning next to Theorem 6 the situation is similar. We again assume $t = r$ and state the following

THEOREM 8. *If $r = t = 4$ and (7.3) holds then*

$$q^2 N(0, \beta) = q^s + K_4(0, \beta) + (q - 1 - J_4)\tau G^s(1). \tag{7.13}$$

If $r = t = 3$ and (7.6) holds then

$$q^2 N(0, \beta) = q^s + K_3(0, \beta) J_3 \psi(\beta)\tau G^{s+1}(1). \tag{7.14}$$

Here K_4, K_3 are determined by (5.2), J_4, J_3 by (7.4), (7.5), (7.7).

8. Conjecture of Weil. Let $N^{(m)}$ denote the number of solutions $\xi_j^{(i)}$ in the extended field $GF(q^m)$ of the system

$$Q_1(\xi^{(1)}) + \cdots + Q_r(\xi^{(r)}) = 0, \qquad \beta_1 Q_1(\xi^{(1)}) + \cdots \beta_r Q_r(\xi^{(r)}) = 0,$$

and let N_m denote the number of "rational points," so that

$$N^{(m)} = 1 + (q^m - 1)N_m. \tag{8.1}$$

Consider first the case covered by Theorem 2 with all s_i even. For brevity we put

$$s_i = 2z_i, \quad s = 2z, \quad \xi_i = \psi((-1)^{z_i}\delta_i)\xi, \quad \xi = \psi((-1)^z)\tau, \tag{8.2}$$

Then (3.5) gives

$$N^{(m)} = q^{2m(z-1)} + q^{m(z-2)}(q^m - 1)\Big(\sum_{i=1}^{r} \xi_i^m q^{mz_i} + \xi^m(q^m + 1 - r)\Big)$$

and by (8.1) this implies

$$N_m = (q^{2m(z-1)} - 1)/(q^m - 1) + q^{m(z-2)}\Big(\sum_{i=1}^{r} \xi_1^m q^{mz_i} + \xi^m(q^m + 1 - r)\Big).$$

Next following Weil [7], p. 507, we put

$$\sum_{m=1}^{\infty} N_m u^{m-1} = d \log Z(u)/du. \tag{8.3}$$

Now a straightforward computation yields

$$\sum_{m=1}^{\infty} N_m u^{m-1} = \sum_{i=0}^{2z-3} q^i/(1-q^iu) + \sum_{i=1}^{r} \xi_i q^{z+z_i-2}/(1-\xi_i q^{z+z_i-2})$$
$$+ \xi q^{z-1}/(1-\xi q^{z-1}u) - (r-1)\xi q^{z-2}/(1-\xi q^{z-2}u);$$

thus (8.3) implies

$$(8.4) \qquad Z(u) = \frac{(1-\xi q^{z-2}u)^{r-1}}{(1-\xi q^{z-1}u)\prod_{i=0}^{2z-3}(1-q^iu)\prod_{i=1}^{r}(1-\xi_i q^{z+z_i-2}u)}.$$

While the right member is of the form conjectured by Weil, $Z(u)$ apparently does not satisfy the conjectured functional equation. However it should be observed that the variety defined by the pair of quadratics has singular points and is accordingly not covered by Weil's conjecture.

In the next place we consider (5.3) with $t = r$. We suppose first that t is even. Then it follows from (5.3) and (5.4) that

$$N^{(m)} = q^{m(s-2)} + (q^m-1)\{1 + \sum_{\lambda \in GF(q^m)} T^{(m)}(\lambda)\}\tau^m G^{ms}(1)q^{-2m},$$

where $T^{(m)}(\lambda)$ indicates that the T-function is defined over $GF(q^m)$. Using (8.1) we get

$$(8.5) \quad N_m = (q^{2m(z-1)}-1)/(q^m-1) + \{1 + \sum_{\lambda} T(\lambda)\}\tau^m G^{2mz}(1)q^{-2m} \qquad (s = 2z).$$

Now by (2.6)

$$(8.6) \qquad \tau^m G^{2mz}(1) = \psi^m((-1)^z\delta_t \cdots \delta_r)q^{mz} = \xi^m q^{mz}.$$

Also Davenport has proved [2], that

$$(8.7) \qquad -\sum_{\lambda \in GF(q^m)} T^{(m)}(\lambda) = q^{m\omega_1} + \cdots + q^{m\omega_{s-1}},$$

where the ω's are the zeros of a certain L-function. Hence using (8.6) and (8.7); (8.5) becomes

$$\sum_{u=1}^{\infty} N_m u^{m-1} = \sum_{i=0}^{2z-3} q^i/(1-q^iu) + \xi q^{z-2}/(1-\xi q^{z-2}u)$$
$$- \sum_{i=1}^{2z-1} \xi q^{z-2+\omega_i}/(1-\xi q^{z-2+\omega_i}u).$$

Therefore by (8.3) we get

$$(8.8) \qquad Z(u) = \prod_{i=1}^{2z-1}(1-\xi q^{z-2+\omega_i}u)/\{(1-\xi q^{z-2}u)\prod_{i=0}^{2z-3}(1-q^iu)\},$$

which is of the conjectured form.

Finally in the case $t = r$, t odd, it follows from (5. 3) and (5. 4) that

$$N^{(m)} = q^{m(s-2)} + (q^m - 1)q^{-2m} \sum_{i=1}^{t} \psi^m((-1)^{\frac{1}{2}(s-s_i)})\tau_i^m q^{\frac{1}{2}m(s+s_i)}. \tag{8. 9}$$

Now put $s = 2z + 1$, $s_i = 2z_i + 1$, $\xi_i = \psi((-1)^{z+z_i})\tau_i$; then (8. 9) implies

$$N_m = (q^{m(2z-1)} - 1)/(q^m - 1) + \sum_{i=1}^{t} \xi_i^m q^{m(z+z_i)},$$

$$\sum_{m=1}^{\infty} N_m u^{m-1} = \sum_{i=0}^{2z-2} q^i/(1 - q^i u) + \sum_{i=1}^{t} \xi_i q^{z+z_i}/(1 - \xi_i q^{z+z_i} u).$$

Hence by (8. 3)

$$Z(u) = \prod_{i=0}^{2z-2} (1 - q^i u)^{-1} \prod_{i=1}^{t} (1 - \xi_i q^{z+z_i} u)^{-1}, \tag{8. 10}$$

which again is of the conjectured form.

9. The case $q = 2^n$. The typical quadratic form is now taken to be

$$Q(u) = \sum_{1 \leq i \leq j \leq r} \alpha_{ij} u_i u_j \qquad (\alpha_{ij} \varepsilon GF(2^n)). \tag{9. 1}$$

Assume that Q is not equivalent to a form in fewer than r indeterminates. Then it is known ([3], § 199) that Q can be reduced to

$$u_1 u_2 + \cdots + u_{2s-1} u_{2s} + u_{2s+1}^2 \qquad (r = 2s + 1), \tag{9. 2}$$

or to

$$u_1 u_2 + \cdots + u_{2s-3} u_{2s-2} + u_{2s-1}^2 + u_{2s-1} u_{2s} + \delta u_{2s}^2 \qquad (r = 2s), \tag{9. 3}$$

where $\delta = 0$ or is such that the form $u^2 + uv + \beta v^2$ is irreducible over $GF(q)$. We may call δ the invariant and r the rank of Q.

We next define

$$G(Q) = \sum e(Q(\xi_1, \cdots, \xi_r)) \tag{9. 4}$$

exactly as in (2. 8). Since

$$\sum_{\xi,\eta} e(\xi\eta) = q, \qquad \sum_{\xi} e(\xi^2) = 0, \tag{9. 5}$$

it follows at once that, according as $r = 2s + 1$ or $r = 2s$.

$$G(Q) = 0 \quad \text{or} \tag{9. 6}$$

$$G(Q) = q^{s-1} \sum_{\xi,\eta} e(\xi^2 + \xi\eta + \delta\eta^2), \tag{9. 7}$$

where δ has the same meaning as in (9. 3). When $\delta = 0$, the sum on the right is evaluated by the first of (9. 5). When $\delta \neq 0$, it is readily seen that the equation

$$\xi^2 + \xi\eta + \delta\eta^2 = \alpha \qquad (\alpha \neq 0) \tag{9. 8}$$

has precisely $q+1$ solutions. Indeed it is clear that the number of solutions of (9.8) is equal to the number of solutions of $\xi^2+\xi\eta+\delta\eta^2=\alpha\beta^2$ $(\alpha\beta\neq 0)$. But since every non-zero number of $GF(2^n)$ can be expressed in the form $\alpha\beta^2$, the assertion about (9.8) follows at once. In other words we have

$$N(\xi^2+\xi\eta+\delta\eta^2=\alpha)=k(\alpha)+2 \qquad (\delta\neq 0), \tag{9.9}$$

where $k(\alpha)$ has the same meaning as in (1.2). Thus (9.7) becomes

$$G(Q)=-q^s \quad (\delta\neq 0), \qquad G(Q)=q^s \quad (\delta=0) \qquad (r=2s). \tag{9.10}$$

It now is easy to show that the number of solutions of $Q(\xi_1,\cdots,\xi_r)=\alpha$ is given by

$$N=q^{2s} \quad (r=2s+1), \qquad N=q^{2s-1}+q^{s-1}k(\alpha)l(\delta) \quad (r=2s), \tag{9.11}$$

where $l(0)=1$, $l(\delta)=-1$ for $\delta\neq 0$. The first half of (9.11) is immediate; to prove the second half we have

$$qN=\sum_{\lambda} e(-\lambda\alpha)G(\lambda Q)=q^{2s}+\sum_{\lambda\neq 0} e(-\lambda\alpha)G(\lambda Q).$$

The stated result now follows on using (9.10).

In the next place we set up a problem like (1.3). Let $N(\alpha,\beta)$ denote the number of solutions of

$$Q_1(\xi^{(1)})+\cdots+Q_r(\xi^{(r)})=\alpha, \qquad \beta_1Q_1(\xi^{(1)})+\cdots+\beta_rQ_r(\xi^{(r)})=\beta, \tag{9.12}$$

where now the Q_j are of the form (9.1). As before Q_j is of rank s_j so that the total number of unknowns is again $s=s_1+\cdots+s_r$. Clearly (2.11) remains valid. We break up the sum as before. Thus the term $\lambda=\mu=0$ again contributes q^s. Secondly if $\lambda=\mu\beta_i$ for fixed i and $\mu\neq 0$ we get

$$q^{s_i}\sum_{\mu\neq 0} e((\beta-\alpha\beta_i)\mu)\prod_{j=1}^{r}{}' G\{(\beta_i-\beta_j)\mu Q_j\} \qquad (j\neq i). \tag{9.13}$$

The remaining terms of (2.11) yield

$$\sum_{\lambda,\mu}{}' e(-\lambda\alpha+\mu\beta)\prod_{i=1}^{r} G\{(\lambda-\mu\beta_i)Q_j\} \qquad (\lambda\neq\mu\beta_i). \tag{9.14}$$

Suppose now that all s_j are even and let δ_j denote the invariant of Q_j. Applying (9.10) to (9.14) we get

$$\sum_{\lambda,\mu}{}' e(-\lambda\alpha+\mu\beta)\cdot q^{\frac{1}{2}s}\prod_{i=1}^{r} l(\delta_i) \qquad (\lambda\neq\mu\beta_i); \tag{9.15}$$

similarly (9.13) yields

$$q^{s_i}\sum_{\mu\neq 0} e((\beta-\alpha\beta_i)\mu)\cdot q^{\frac{1}{2}(s-s_i)}\prod_{j=1}^{r}{}' l(\delta_i) \qquad (j\neq i). \tag{9.16}$$

Hence we obtain

$$q^2N(\alpha,\beta) = q^s + lq^{\frac{1}{2}s}\sum_{i=1}^{r}(q^{\frac{1}{2}s_i}l(\delta_i) - 1)k(\beta - \alpha\beta_i) + lq^{\frac{1}{2}s}(k(\alpha)k(\beta) + k(\alpha) + k(\beta)), \tag{9.17}$$

where $l = \prod_{i=1}^{r} l(\delta_i)$. In particular for $\alpha = \beta = 0$, (9.17) becomes

$$q^2N(0,0) = q^s + (q-1)lq^{\frac{1}{2}s}\{\sum_{i=1}^{r} q^{\frac{1}{2}s_i}l(\delta_i) + q + 1 - r\}. \tag{9.18}$$

As for the general case $s_1, \cdots, s_t$ odd, $s_{t+1}, \cdots, s_r$ even, it is clear from (9.6) that (9.14) vanishes for $t \geqq 1$; indeed the same is true of (9.13) for $t \geqq 2$. This evidently implies

$$N(\alpha,\beta) = q^{s-2} \qquad (t \geqq 2). \tag{9.19}$$

In the remaining case ($t = 1$) we get

$$q^2N(\alpha,\beta) = q^s + q^{\frac{1}{2}(s+s_1)}k(\beta - \alpha\beta_1)\prod_{j=2}^{r} l(\delta_j) \qquad (t = 1). \tag{9.20}$$

We may now state

THEOREM 9 ($q = 2^n$). *Let $s_1, \cdots, s_t$ be odd, $s_{t+1}, \cdots, s_r$ even. Then the number of solutions of* (9.12) *is given* (9.17) *for* $t = 0$, (9.19) *for* $t \geqq 2$, (9.20) *for* $t = 1$.

It may be remarked that it is easy to give a direct proof of (9.19) and (9.20) using only (9.2) and (9.11). For to begin with the number of solutions of (9.12) is clearly the same as the number of solutions of the system

$$\begin{aligned} &Q_1(\xi^{(1)}) + \cdots + Q_r(\xi^{(r)}) = \alpha, \\ &(\beta_2 - \beta_1)Q_2(\xi^{(2)}) + \cdots + (\beta_r - \beta_1)Q_r(\xi^{r}) = \beta - \alpha\beta_1, \end{aligned} \tag{9.21}$$

Now the number of solutions of the second of (9.21) can be obtained readily from (9.11). Finally we observe that for fixed $\xi^{(2)}, \cdots, \xi^{(r)}$, the number of solutions of the first of (9.21) is $q^{\frac{1}{2}(s_1-1)}$. Formulas (9.19) and (9.20) now follow.

DUKE UNIVERSITY.

REFERENCES.

[1] L. Carlitz, "The singular series for sums of squares of polynomials," *Duke Mathematical Journal*, vol. 14 (1947), pp. 1105-1120.

[2] H. Davenport, "On character sums in finite fields," *Acta Mathematica*, vol. 71 (1939), pp. 99-121.

[3] L. E. Dickson, *Linear groups*, Leipzig, 1901.

[4] H. Hopf, "Über die Verteilung quadratischer Reste," *Mathematische Zeitschrift*, vol. 32 (1930), pp. 222-231.

[5] E. Jacobsthal, "Über die Darstellung der Primzahlen der Form $4n + 1$ als Summe zweier Quadrate," *Journal für die reine und angewandte Mathematik*, vol. 132 (1907), pp. 288-245.

[6] A. Weil, "On the Riemann hypothesis in function fields," *Proceedings of the National Academy of Sciences*, vol. 27 (1941), pp. 345-347.

[7] ———, "Numbers of solutions of equations in finite fields," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 497-508.

[8] A. L. Whiteman, "Cyclotomy and Jacobsthal sums," *American Journal of Mathematics*, vol. 74 (1952), pp. 89-99.

[9] ———, "Theorems analogous to Jacobsthal's theorem," *Duke Mathematical Journal*, vol. 16 (1949), pp. 619-626.

TOPOLOGICAL PROPERTIES OF HOMOMORPHISMS BETWEEN BANACH ALGEBRAS.*[1]

By BERTRAM YOOD.

1. Introduction. Let B and B_1 be two (real or complex) Banach algebras. A number of known results [3], [6], [12], [13] give conditions on the algebras under which certain classes of homomorphisms of B into B_1 automatically possess topological properties. These results are useful in representation theory, in the uniqueness of norm problem [12] and are of intrinsic interest. In §§ 3, 4 we give a unified treatment of results of this type and related matters where we have tried to reduce the assumptions on the algebras to a minimum. Here, in algebras with an involution, the involution can be either an isomorphism or an anti-isomorphism.[2] The basic type of Banach algebra studied is the ρ^*-algebra where all that is assumed is that it is possible to find an involution and a non-negative sub-additive function $|x|$ defined on B, vanishing only for $x = 0$ and majorized by the spectral radius for self-adjoint and skew elements. This is a generalization of the A^*-algebras of Rickart [12]. § 3 is devoted to showing that all of his results for these algebras carry over to the ρ^*-algebra case. In particular every ρ^*-algebra has the uniqueness of norm property.

For the bi-continuity result (§ 4) the domain algebra B is a ρ^*-algebra where $|x - y|$ is a metric making B a space of type (F) and where self-adjoint elements have real spectra. This is a generalization of B^*-algebras and the *-algebras of Arens. Any *-isomorphism of B into a ρ^*-algebra is bi-continuous.

Every ρ^*-algebra is semi-simple. In § 5 we pass to the study of algebras with non-trivial radical in the case of a complex commutative regular Banach algebra B. We obtain results on homomorphisms (Theorems 5.3 and 5.4) which, when B is semi-simple or is a B^*-algebra, in the main reduce to recent results of Rickart [13] and Kaplansky [6].

---

* Received November 28, 1952; substantially revised, August 24, 1953.

[1] This research was supported in part by the United States Air Force under Contract No. AF 18(600)-685 monitored by the Office of Scientific Research.

The author is indebted to the referee for his helpful suggestions.

[2] This is possible, in part, because the present treatment unlike its predecessors does not use the combination xx^* if x is not normal. Also note that $xx^* = 0$ is possible without $x = 0$.

This paper is a continuation and extension of the work of Rickart and Kaplansky [6], [12], [13]; our scientific debt to these authors is cheerfully acknowledged.

2. Notation. Let B be a real Banach algebra. For the notion of spectrum see [6], [12]. $\alpha(x)$ denotes the spectrum of $x \in B$ which is defined to be the usual complex algebra spectrum of $(x, 0)$ in the complexification of B. For our purposes we may omit 0 from $\alpha(x)$. If E is a subalgebra of B we use $\alpha(x \mid E)$ for the spectrum of x as an element of E. $\rho(x)$ is the spectral radius of x $(= \sup \mid \lambda \mid, \lambda \in \alpha(x))$. If B admits complex scalars then $\alpha(x)$ can differ from the usual spectrum $\sigma(x)$ of x. Some of our hypotheses involve the spectrum as a point-set, but our results are also valid if B is a complex Banach algebra and the term spectrum refers instead to $\sigma(x)$ (even though the involution and homomorphisms need only be real linear). Essentially the same arguments prevail. By an *involution* on B is meant a mapping $x \to x^*$ which is additive, real-homogeneous, of period two and either $(xy)^* = y^*x^*$ for every $x, y \in B$ or $(xy)^* = x^*y^*$ for every x, y. Let $H(K)$ be the set of self-adjoint (skew) elements of B. Then, as usual, B is the direct sum $H \oplus K$. B is a *ρ^*-algebra* if there exists an involution and a real function $\mid x \mid$ on B, $\mid x \mid \geqq 0$, $\mid x \mid = 0$ if and only if $x = 0$, $\mid x + y \mid \leqq \mid x \mid + \mid y \mid$ and $\mid x \mid \leqq \rho(x)$, $x \in H \cup K$. Any such function $\mid x \mid$ is called an *auxiliary function* for B. If $\mid x \mid$ can be chosen so that $\mid x - y \mid$ is a metric in which B is a complete space of type (F); ([2], p. 35), B is a *ρ_c^*-algebra*. B is *Hermitian real* if the spectrum of every $x \in H$ is real and C-spectral if, in addition, the spectrum of every $x \in K$ is pure-imaginary.[3]

3. ρ^*-algebras. By an A^*-algebra [12] is meant a Banach algebra in which there is an (anti-isomorphic) involution $x \to x^*$ and a norm $\mid x \mid$ in which it is a normed algebra and a number $k > 0$ such that $\mid x \mid^2 \leqq k \mid x^*x \mid$ for all x. By the arguments of [12], p. 626, every A^*-algebra is a ρ^*-algebra. The converse is an open question but appears to us to be highly unlikely, there being no evidence in its favor. Note also that the definition of a ρ^*-algebra permits involutions not allowed in the A^*-case. A difficulty in trying to exhibit a ρ^*-algebra which is not an A^*-algebra is that somehow one must examine all possible anti-isomorphic involutions and with each all possible normed algebra norms. We feel that the notion of a ρ^*-algebra is

[3] The C-spectral property does not follow from the Hermitian-real property for real Banach algebras. Consider, for example, the algebra of all real continuous functions on $[0, 1]$ where $x^*(t) = x(1 - t)$.

simpler and more natural than that of an A^*-algebra. In any case we shall develope the theory of ρ^*-algebras up to the point reached by A^*-algebras.

3.1. LEMMA.[4] *Let B be a Banach algebra with an involution. Then the following statements are equivalent.*

(1) *The involution is continuous.*

(2) *H and K are closed.*

(3) *Suppose that for $x = u + v$ where $u \in H$, $v \in K$ we define a norm $\| x \|_1 = 2(\| u \| + \| v \|)$. In this norm B is a Banach algebra.*

Assume (1). Let T be the involution and I the identity on B. Since $I \pm T$ is continuous, (2) follows.

To show that the norm $\| x \|_1$ defined in (3) makes B a normed algebra it is sufficient to verify that $\| xy \|_1 \leqq \| x \|_1 \| y \|_1$. Let $x = u_1 + v_1$, $y = u_2 + v_2$ where each $u_k \in H$, $v_k \in K$. Then

$$xy = u_1u_2 + v_1v_2 + (u_1v_2 + v_1u_2) = w,$$

say, and

$$w^* = u_2u_1 + v_2v_1 - (v_2u_1 + u_2v_1).$$

The decomposition of w according to $B = H \oplus K$ gives

$$w = \tfrac{1}{2}(w + w^*) + \tfrac{1}{2}(w - w^*)$$

and

$$\begin{aligned} \| w \|_1 = \| w + w^* \| + \| w - w^* \| &\leqq 2(\| w \| + \| w^* \|) \\ &\leqq 4(\| u_1 \| + \| v_1 \|)(\| u_2 \| + \| v_2 \|) = \| x \|_1 \| y \|_1. \end{aligned}$$

Now assume (2). Let $\| x_m - x_n \|_1 \to 0$ and write $x_n = u_n + v_n$ with $u_n \in H$, $v_n \in K$. Then $\| u_m - u_n \| \to 0$ and $\| v_m - v_n \| \to 0$. Consider $u, v \in B$ where $u_n \to u$, $v_n \to v$. Write $u = r + s$ with $r \in H$, $s \in K$. Then $u_n - r \in H$ and $\| u_n - r - s \| \to 0$. By (2), $s = 0$ and thus $u \in H$. Similarly, $v \in K$. Hence for $x = u + v$, $\| x_n - x \|_1 \to 0$ and B is complete in the norm of (3).

Assume (3). Since $\| x \|_1 \geqq \| x \|$, by [2], p. 41, the two norms give the same topology. Since $\| x^* \|_1 = \| x \|_1$, the involution is continuous in the original norm.

3.2. LEMMA. *Let B be a Banach algebra with an involution. If $H(K)$ is closed then, for y in $H \cap K'$ $(K \cap H')$, $y^2 = 0$.*

Clearly that $H(K)$ be closed is equivalent to $K \cap H'$ $(H \cap K')$ being the zero element. Let $H(K)$ be closed, $x_n \in K(H)$, $x_n \to y$, $y \in H(K)$.

[4] The modification of the proof to cover the case where the involution is an isomorphism is clear.

Let $v \in K$. Then $z_n = vx_n + x_n v \in H(K)$ and $z_n \to vy + yv \in K(H)$. Thus, by the closure assumption, $vy + yv = 0$. Then $vy + yv = 0$ also for all $v \in K'$ and we obtain $y^2 = 0$.

3.3. LEMMA. *Let B be a Banach algebra with an involution where there is a non-negative sub-additive function $|x|$ defined on H, vanishing only for $x = 0$ such that $|x| \leqq \rho(x)$, $x \in H$. Then $y^2 = 0$ for each $y \in K \cap H'$.*

Suppose that $x_n \in H$, $x_n \to y$, $y \in K$. Now we may suppose that $|-x| = |x|$ for we could use $(|x| + |-x|)/2$. We have

$$(1) \qquad |(|x| - |y|)| \leqq |x - y| \leqq \|x - y\|, \qquad (x, y \in H).$$

Let $\{w_n\}$ be a Cauchy sequence in H. By (1), $\{|w_n|\}$ is a Cauchy sequence of real numbers. Define $f(w)$ on the closure of H by the rule that if $w = \lim w_n$, $w_n \in H$, then $f(w) = \lim |w_n|$. If also $w = \lim y_n$, $y_n \in H$, then, by (1), $\lim |w_n| = \lim |y_n|$ so that $f(w)$ is well-defined. Clearly $f(w)$ is sub-additive on its domain of definition. A theorem of Hille ([5], Theorem 22.9.1) asserts that if G is any open set in the complex plane containing $\alpha(w)$ and 0, then there exists $\epsilon > 0$ such that $\alpha(w_n) \subset G$ for all n where $\|w_n - w\| < \epsilon$ (see also [10]). From this we obtain

$$(2) \qquad f(w) = \lim |w_n| \leqq \liminf \rho(w_n) \leqq \rho(w) \leqq \|w\|,$$

$$(3) \qquad f(w) \leqq \rho(w) = \rho(w^*) \leqq \|w^*\|.$$

Since H is closed under the operation of squaring, so is its closure and thus $(y \pm x_n)^2$ are in the domain of definition of f. Also $y^2 \in H$. By (2) and (3),

$$\begin{aligned} 0 \leqq |(y^2 + x_n^2)/2| &\leqq f[(y + x_n)^2/4] + f[(y - x_n)^2/4] \\ &\leqq \{\|[(y + x_n)^2]^*\| + \|(y - x_n)^2\|\}/4 \\ &\leqq [\|y - x_n\|^2]/2 \to 0. \end{aligned}$$

Also since $\|y^2 - x_n^2\| \to 0$ then $|(y^2 - x_n^2)/2| \to 0$. Since

$$|y^2| \leqq |(y^2 + x_n^2)/2| + |(y^2 - x_n^2)/2|$$

we have $|y^2| = 0$ and thus $y^2 = 0$.

3.4. THEOREM. *Let B be as in Lemma 3.3. If $x^2 = 0$ implies $x = 0$ for $x \in (H \cap K') \cup (K \cap H')$ then the involution in B is continuous. The involution in a ρ^*-algebra is continuous.*

By Lemma 3.3 and hypothesis, $K \cap H'$ reduces to 0 and H is closed. With the aid of Lemma 3.2, our hypotheses now show also that $H \cap K' = 0$ and K is closed. By Lemma 3.1, the involution in B is continuous.

If B is a ρ^*-algebra, then $x \in H \cup K$, $x^2 = 0$ implies that $\rho(x) = |x| = 0$ and $x = 0$.

3.5. LEMMA. *Let A be a real algebra with an involution where no element x in $H \cup K$, $x \neq 0$ is quasi-nilpotent. Then A is semi-simple.*

By ([5], Lemma 22.15.1), either $A = R$, the radical of A, or R is the intersection of the maximal regular left (right) ideals of A. Consider the latter case. If the involution $x \to x^*$ is an anti-isomorphism then it takes a maximal regular left (right) ideal of A onto a maximal regular right (left) ideal of A and each maximal regular left (right) ideal is the image of such a right (left) ideal under the involution. Then in all situations the involution takes R onto R. Hence for any $y \in R$, $y \pm y^* \in R$. Since these are quasi-nilpotent, our hypotheses give $y = 0$.

Clearly any *-subalgebra (closed or not) of a ρ^*-algebra satisfies the hypotheses of this lemma.

3.6. LEMMA. *Let B be a ρ^*-algebra and let B_1 denote a Banach algebra which is a *-subalgebra of B. Then there exists a constant C such that $\|x\| \leqq C\|x\|_1$ for $x \in B_1$, where $\|x\|$, $\|x\|_1$ are the norms in B, B_1 respectively.*

Since $\rho(x \mid B_1) \geqq \rho(x \mid B)$ for x in B_1 it is clear that B_1 is a ρ^*-algebra with the same auxiliary function $|x|$ as B. By Lemma 3.2 the involution is continuous in each of B and B_1. Let $x \in B$, $x = u + v$, $u \in H$, $v \in K$, Then $|x| \leqq |u| + |v| \leqq \rho(u) + \rho(v) \leqq \|u\| + \|v\|$. Since, by Lemma 3.1, $2(\|u\| + \|v\|)$ is a complete norm we obtain for some $c_1 > 0$, $|x| \leqq c_1\|x\|$, $x \in B$. Similarly for some $c_2 > 0$, $|x| \leqq c_2\|x\|_1$, $x \in B_1$. Now $|x|$ is a sub-additive function on B vanishing only for $x = 0$. As in the proof of [12], Theorem 5.4, we obtain the conclusion.

3.7. THEOREM. *Every ρ^*-algebra has the uniqueness of norm property. Let T be a homomorphism of a Banach algebra B into a ρ^*-algebra B_1 where $T(B) = [T(B)]^*$. Then T is continuous.*

With the above results in hand, the proof is the same as that used by Rickart ([12], p. 627).

In this connection we note that the uniqueness of norm property for a Banach algebra B implies that every possible involution on B is continuous. This follows by an argument of Rickart ([11], Theorem 3.1).

3.8. COROLLARY. *Let T be a homomorphism of a Banach algebra B into a Banach algebra B_1 where there exists an auxiliary function $|x|$ with*

$|x| \leqq \rho(x)$, $x \in B_1$. *Then T is continuous. In particular, this holds if B_1 is a BQ^*-algebra.*

The continuity of T follows directly from Theorem 3.7 since it is permissable to use the identity as the involution in B_1 even in the non-commutative case. If B_1 is a BQ^*-algebra [1] then results of Arens shows that $\rho(x) = \| x \|$, $x \in B_1$. This corollary is an extension of [12], Cor. 6, which gives the continuity of T where B_1 is a semi-simple commutative Banach algebra.

4. Isomorphisms and ρ_c^*-algebras. This section deals with the properties of an algebraic isomorphism T defined on a Hermitian-real ρ_c^*-algebra. Conditions are given which establish the continuity of T^{-1} and the bi-continuity of T.

Consider the notion of C-symmetry [6]. Since this refers to the behavior of self-adjoint *commutative* subalgebras the notion makes sense in our extended meaning of the term involution.

4.1. THEOREM.

(1) *A Banach algebra B is a ρ_c^*-algebra if and only if there exists an involution and $k > 0$ where $\| x \| \leqq k\rho(x)$, $x \in H \cup K$.*

(2) *The class of all commutative ρ_c^*-algebras coincides with the class of all commutative Banach algebras for which $\rho(x)$ is a complete norm.*

(3) *If B is a C-spectral ρ_c^*-algebra then B is C-symmetric.*

(4) *If B of* (3) *admits complex scalars and is commutative then it is a B^*-algebra under an equivalent norm.*

(1). Let B be a ρ_c^*-algebra with auxiliary function $|x|$. By Lemma 3.1 and Theorem 3.4 there exists $M > 0$ such that $|x| \leqq M \| x \|$, $x \in B$. Hence, by [2], p. 41, if $|x_n| \to 0$ then $\| x_n \| \to 0$. Suppose that the inequality in (1) is invalid for each $k > 0$. Then there exists a sequence $\{x_n\}$, $x_n \in H \cup K$, $\| x_n \| = 1$, $1 > n\rho(x_n)$. Thus $|x_n| \leqq \rho(x_n) < n^{-1}$ and $|x_n| \to 0$. This is a contradiction.

(2). Let B be a commutative ρ_c^*-algebra. By commutativity and (1),

$$\| x \| \leqq \| (x + x^*)/2 \| + \| (x - x^*)/2 \|$$
$$\leqq k[\rho(x + x^*) + \rho(x - x^*)]/2 \leqq 2k\rho(x) \leqq 2k \| x \|.$$

If B has $\rho(x)$ as a complete norm then clearly B is a ρ_c^*-algebra in every involution including the identity.

(3). Let E be any closed self-adjoint commutative subalgebra. Then E is contained in a maximal self-adjoint commutative subalgebra A which is closed. C-symmetry for E follows if we show it for A. Let A_1 be the complexification of A ([6], [12]) and $\mathfrak{M}$ the locally compact space of maximal regular ideals of A_1. A can be considered as imbedded in A_1. Now for $x \in A$, $\alpha(x \mid A_1) = \alpha(x \mid A) = \alpha(x \mid B)$. If x is self-adjoint (skew) in A, its spectrum there is real (pure imaginary). If $x = u + v$, $u \in H$, $v \in K$, then, under the Gelfand representation of A_1, $x^*(M) = u(M) - v(M) = \overline{x(M)}$, $M \in \mathfrak{M}$. Since $\rho(xx^* + yy^*) \geqq \rho(xx^*) = \rho(x^2) = \rho(x)\rho(x^*)$, for $x, y \in A$, A is C-symmetric by [6], Theorem 4. 6, and (2).

(4). By (3) and [6], Theorem 4. 5, the involution is conjugate-linear. (4) follows from the formulas of (3).

4. 2. THEOREM. *Let B be a Hermitian-real Banach algebra where each closed subalgebra E generated by a self-adjoint element is a Banach space in the spectral radius norm $\rho(y \mid E)$. Let T be an algebraic isomorphism of B into a Banach algebra. Then $\rho(T(x)) = \rho(x)$, $x \in H \cup K$.*

By Theorem 4. 1 the hypothesis on E is automatically satisfied if B is a ρ_c*-algebra.

Let E_0 be the smallest subalgebra generated by the self-adjoint element u and let E be its closure. For $x \in E_0$ the reality of $\alpha(x \mid B)$ implies that $\alpha(x \mid E) = \alpha(x \mid B)$ (see the arguments of [6], Theorem 4. 2). However for any $x \in E$, $\alpha(x \mid E)$ is real (as can be seen from the Gelfand representation of the complexification of E) and again $\alpha(x \mid E) = \alpha(x \mid B)$. Thus $\rho(x \mid E) = \rho(x)$, $x \in E$. By hypothesis there exists $k > 0$ such that $\| x \| \leqq k\rho(x) \leqq k \| x \|$ on E. The reality of the spectra implies that if M is any maximal regular ideal in E, E/M is isomorphic to the real number system. The Gelfand representation $x \to x(M)$ takes E into $C_r(\mathfrak{M})$, the algebra of all real continuous functions on $\mathfrak{M}$ which vanish at infinity, where $\mathfrak{M}$ is the space of maximal regular ideals of E. By the Stone-Weierstrass theorem, E is isomorphic, as a Banach algebra, to $C_r(\mathfrak{M})$. Therefore E is semi-simple and regular and, by [13], Theorem 1, Cor. 1, $\rho(u) = \rho(T(u))$. Let $v \in K$. Then $v^2 \in H$ and

$$[\rho(v)]^2 = \rho(v^2) = \rho(T(v^2)) = \rho([T(v)]^2) = [\rho(T(v))]^2.$$

4. 3. THEOREM. *Let B be a Hermitian-real ρ_c*-algebra. Let T be an algebraic isomorphism of B into a Banach algebra where for some $c > 0$, $\| T(x^*) \| \leqq c \| T(x) \|$, $x \in B$. Then T^{-1} is continuous on $T(B)$.*

By Theorem 4. 1 there exists $k > 0$ such that $k \| x \| \leqq \rho(x)$, $x \in H \cup K$. Theorem 4. 2 gives

$$\text{(a)} \qquad \| T(x) \| \geqq k \| x \|, \qquad x \in H \cup K.$$

Let $B_1 = \overline{T(B)}$. In $T(B)$ we define the operation $[T(x)]' = T(x^*)$. Clearly this provides a continuous involution on $T(B)$. This involution may be extended by continuity to a continuous mapping V defined on B_1. It is easy to verify that V is an involution on B_1. Let $\| y \|_1$ be the equivalent norm for B_1 given by Lemma 3. 1. By (a), for $x = u + v$, $u \in H$, $v \in K$,

$$\| T(x) \|_1 = 2(\| T(u) \| + \| T(v) \|) \geqq 2k(\| u \| + \| v \|) \geqq k \| x \|.$$

For the case where B is a complex B^*-algebra this result is due to Rickart ([13], Theorem 3, Cor.).

The following example shows that the Hermitian-real hypothesis cannot be dropped entirely even in the commutative case. Let B be the Banach algebra of all complex-valued functions holomorphic for $|z| < 1$ and continuous on $|z| \leqq 1$ in the complex plane with $\| f \| = \sup |f(z)|$, $|z| \leqq 1$. Let E be the set $|z| \leqq 1/2$ and B_1 be the Banach algebra of all continuous complex-valued functions on E. Let the identity be the involution on B and B_1 and let T be the *-isomorphism which sends $f(z) \in B$ into $f(z) \in B_1$, where z is restricted to E. T is obviously continuous and all the hypotheses for the continuity of T^{-1} are satisfied except the Hermitian-real requirement for B. But if we take $f_n(z) = z^n$ we see that $\| f_n \| = 1$ while $\| T(f_n) \| \to 0$.

The next result is the analogue of a well-known theorem on B^*-algebras (see [6], Theorem 6. 4). In our context we must be content with bi-continuity instead of isometry.

4. 4. COROLLARY. *Let B be a Hermitian-real ρ_c*-algebra and B_1 be a ρ*-algebra. Let T be an isomorphism of B into B_1, $T(B) = [T(B)]^*$. Then T is bi-continuous if and only if there exists $M > 0$ such that $\| T(x^*) \| \leqq M \| T(x) \|$, $x \in B$. If T is a *-isomorphism, T is bi-continuous.*

The continuity of T is given by Theorem 3. 7. If the inequality holds we have bi-continuity by Theorem 4. 3, if T is bi-continuous then the inequality follows from Theorem 3. 4.

4. 5. COROLLARY.[5] *Let T be a homomorphism of a B^*-algebra B into a ρ*-algebra B_1 where $[T(B)]^* = T(B)$. Then $T(H)$ and $T(K)$ are closed in B_1.*

[5] Here the involution is an anti-isomorphism.

By Theorem 3. 7, T is continuous and $T^{-1}(0)$ is closed. By [6], Theorem 7. 2, the proposition is reduced to the case where T is one-to-one. Theorem 4. 2 implies that T^{-1} is continuous on $T(H)$ and $T(K)$. If $\| T(x^*)\| \leqq C \| T(x)\|$, $x \in B$, then $T(B)$ is closed by Theorem 4. 3.

4. 6. THEOREM.[6] *Let B be a complex symmetric ρ^*-algebra (ρ_c^*-algebra) with an identity. Then there exists a C^*-algebra B_1 and a continuous (bi-continuous) *-isomorphism T of B into (onto) B_1.*

By [6], p. 402, B is C-spectral and the involution is continuous by Theorem 3. 4. Hence we may, by Lemma 3. 1, use an equivalent norm in which $\| x^* \| = \| x \|$. We now use the results of Gelfand and Neumark as given in [8], [9]. As B is semi-simple (by 3. 5) it follows from [9], p. 87, that B coincides with its reduced ring in the sense of [9], p. 49. That $x^*x = 0$ implies $x = 0$ in B follows from [9], p. 87. By [9], Theorem 1, p. 49, there exists a norm $\| x \|_1$ for B in which it is a normed algebra and such that its completion B' in this norm is a C^*-algebra where the involution in B' is an extension of that of B. Thus the identity mapping T of B into B' is an algebraic *-isomorphism. The conclusions now follow by appealing to [9], p. 33 and Corollary 4. 4.

5. On regular Banach algebras. For the remainder of this paper we adopt the following notation. $B(B_1)$ is a complex commutative Banach algebra with radical $R(R_1)$ and whose space of maximal regular ideals is $\mathfrak{M}(\mathfrak{N})$. Let $\mathfrak{M}'(\mathfrak{N}')$ be the one-point compactification of $\mathfrak{M}(\mathfrak{N})$ by adjoining $B(B_1)$. Let T be an algebraic (complex-homogeneous) homomorphism of B into B_1. Each $N' \in \mathfrak{N}'$ determines by the following rule a unique element $T^*(N')$ of $\mathfrak{M}'$ where

$$x[T^*(N')] = T(x)(N'), \qquad x \in B, \tag{1}$$

(see [3], p. 202; [7], p. 76). If $T(B) \subset N'$ then $T^*(N') = B$, otherwise $T^*(N')$ is the unique member M' of $\mathfrak{M}'$ such that $T(M') \subset N'$. T^* is a continuous mapping of $\mathfrak{N}'$ into $\mathfrak{M}'$ in the Gelfand topologies for $\mathfrak{N}'$ and $\mathfrak{M}'$ so that $T^*(\mathfrak{N}')$ is closed in $\mathfrak{M}'$. In the special case where $T(B)$ is dense in B_1, T^* is a homeomorphism of $\mathfrak{N}(\mathfrak{N}')$, onto a closed set in $\mathfrak{M}(\mathfrak{M}')$.

Following the usage in [7] and elsewhere, for a set $\mathfrak{F}$ in $\mathfrak{M}$ (or $\mathfrak{M}'$) we define the *kernel* of $\mathfrak{F}$, $k(\mathfrak{F}) = \cap M$, $M \in \mathfrak{F}$ and, for an ideal I in B, the *hull* of I in $\mathfrak{M}(\mathfrak{M}')$ is $h(I) = \{M \in \mathfrak{M}(M' \in \mathfrak{M}') \mid M(M') \supset I\}$. *If $T(B)$ is* dense in B_1 and if $\mathfrak{M}$ and $\mathfrak{N}$ are both given the Stone topology (hull-kernel

[6] Here the involution is a conjugate-linear anti-isomorphism.

topology), then it is easy to see that T^* is again continuous on $\mathfrak{N}$. B is called *regular* ([14]; [7], p. 82) if the Stone and Gelfand topologies for $\mathfrak{M}$ agree.

5.1. LEMMA. *If B is regular, then $B_1 = \overline{T(B)}$ is also regular.*

We use (S) or (G) when it is desirable to specify which of the Stone or Gelfand topologies is used. By the above, the mappings $T^*: \mathfrak{N}(S) \to \mathfrak{T}^*(\mathfrak{N})(S) = T^*(N)(G)$ and $T^{*-1}: T^*(\mathfrak{N})(G) \to \mathfrak{N}(G)$ are continuous. Hence the indentity mapping $I: \mathfrak{N}(S) \to \mathfrak{N}(G)$ is continuous. Since the Gelfand topology is stronger than the Stone topology, I is bi-continuous and B_1 is regular.

We next give some elementary properties of T^*.

5.2. LEMMA.

(i) $k(T^*(\mathfrak{N}')) = T^{-1}(R_1)$.

(ii) $T^*(\mathfrak{N}') = \mathfrak{M}'$ *implies* $R = T^{-1}(R_1)$.

(iii) *If B is regular the converse holds.*

(iv) *Let $\mathfrak{N}'_1$ be the compactified space of maximal regular ideals of $\overline{T(B)}$. If $\overline{T(B)}$ is regular then $T^*(\mathfrak{N}'_1) = T^*(\mathfrak{N}')$.*

(i). Note that $x(M') = 0$ for all $M' \in T^*(\mathfrak{N}')$ if and only if $T(x)(N') \equiv 0$ by formula (1). This implies (i) and (ii).

(iii). Suppose that $T^*(\mathfrak{N}') \neq \mathfrak{M}'$. Then by regularity and the closure of $T^*(\mathfrak{N}')$ there exists $M_0 \in \mathfrak{M}$, $M_0 \notin T^*(\mathfrak{N})$ and $x \in B$ such that $x(M_0) = 1$ and $x(T^*(\mathfrak{N})) = 0$. Then $x \in T^{-1}(R_1) = R$ which is a contradiction.

(iv). Let I be the identity mapping of $\overline{T(B)}$ into B_1. Since $R_1 \cap \overline{T(B)}$ is the radical of $\overline{T(B)}$, then $I^*(\mathfrak{N}') = \mathfrak{N}'_1$ by (iii). Clearly both $T^*(\mathfrak{N}')$ and $T^*(\mathfrak{N}'_1)$ contain B. Suppose that $T^*(N'_1) = M' \neq B$, $N'_1 \in \mathfrak{N}'_1$. Let $I^*(N') = N'_1$. Then $T(M') \subset N'$ and $T(B) \not\subset N'$ (as otherwise $\overline{T(B)} \subset N'$ and $I^*(N') = \overline{T(B)}$). Thus $T^*(N') = M'$. Suppose $T^*(N') = M' \neq B$, $N' \in \mathfrak{N}'$. Then $T(M') \subset N' \cap \overline{T(B)}$. But $N' \cap \overline{T(B)} \neq \overline{T(B)}$ and is a maximal regular ideal of $\overline{T(B)}$.

5.3. THEOREM. *Let B be regular. Then the following formulas hold for $T^*(\mathfrak{N}')$.*

(a) $T^*(\mathfrak{N}') = h(T^{-1}(R_1))$.

(b) $T^*(\mathfrak{N}') = h(T^{-1}[\overline{T(R)}])$.

As previously noted, $T^*(\mathfrak{N}')$ is closed in $\mathfrak{M}'$. Thus $T^*(\mathfrak{N}')$ with the point B removed is closed in $\mathfrak{M}$, hence in the Stone topology of $\mathfrak{M}$ by regularity. Then, by Lemma 5.2, $T^*(\mathfrak{N}') = hkT^*(\mathfrak{N}') = h[T^{-1}(R_1)]$.

The proof for formula (b) is largely modeled after [13], Theorem 1. For real Banach algebras Rickart's theorem gives (b) if B is semi-simple and T is 1-1. Set $J = T^{-1}[\overline{T(R)}]$ and $\mathfrak{F} = h(J)$. Incidentally J is an ideal in B. Since $T(R) \subset R_1$ and R_1 is closed, $T^{-1}(R_1) \supset J$ and then, by (a), $\mathfrak{F} \supset T^*(\mathfrak{N}')$. Suppose that there exists $M_0 \in \mathfrak{F}$, $M_0 \notin T^*(\mathfrak{N}')$. Then $M_0 \neq B$ and, since $T^*(\mathfrak{N}')$ is compact in $\mathfrak{M}'$, there exists an open set $\mathfrak{G}$ containing M_0 whose closure is disjoint with $T^*(\mathfrak{N}')$ and an open set $\mathfrak{G}_1$ in $\mathfrak{M}'$ such that $M_0 \in \mathfrak{G}_1$ and $\bar{\mathfrak{G}}_1 \subset \mathfrak{G}$. As in the reference cited, a result of Šilov ([14], p. 37), gives the existence of $u, v \in B$, such that $u(M') = 1$, $M' \in \bar{\mathfrak{G}}_1$, $u(M') = 0$, $M' \notin \mathfrak{G}$, $v(M') = 1$, $M' \in \bar{\mathfrak{G}}$, $v(M') = 0$, $M' \in T^*(\mathfrak{N}')$ and $uv(M') = u(M')$, $M' \in \mathfrak{M}'$. Then [7] $u \circ v(M') = v(M')$ and $u \circ v = v + z$ where $z \in R$. Since $T(v)(N') \equiv 0$, $T(v)$ lies in the radical of $\overline{T(B)}$. Hence there exists $w \in \overline{T(B)}$ such that $T(v) \circ w = 0$. Let $w_n = T(v_n)$, $v_n \in B$, $w_n \to w$. Then

$$T(u) = T(u) \circ [T(v) \circ w] = [T(u \circ v)] \circ w = [T(v + z)] \circ w$$
$$= T(v) \circ w + T(z) - T(z)w = T(z) - T(z)w = \lim T(z - zv_n).$$

But $z - zv_n \in R$. Hence $T(u) \in \overline{T(R)}$ which implies that u is in every element of $\mathfrak{F}$. However $u(M_0) = 1$, $M_0 \in \mathfrak{F}$ which is a contradiction.

An example shows that the requirement that B be regular cannot be dropped. Consider the example utilized at the end of Theorem 4. 3. In the notation used there B is not regular since its space of maximal ideals is not a Hausdorff space in the Stone topology [4]. B and B_1 are semi-simple and $h(T^{-1}(R_1)) = \mathfrak{M}$. However, by the fact [4] that every maximal ideal of B is of the form $\{f \in B \mid f(z_0) = 0\}$, it can be seen that $T^*(\mathfrak{N})$ is the set of maximal ideals corresponding to the complex numbers $|z| \leqq 1/2$.

We make the following definitions. Let $\mathfrak{F}$ be any subset of $\mathfrak{M}'$ and $C(\mathfrak{F})$ be the algebra of all bounded complex continuous functions on $\mathfrak{F}$ with the usual norm. For $x \in B$ let $x(M' \mid \mathfrak{F})$ be the function $x(M')$ restricted to $\mathfrak{F}$. If the set of functions $x(M' \mid \mathfrak{F})$, $x \in B$ is closed in $C(\mathfrak{F})$ we say that B is *functionally closed with respect to* $\mathfrak{F}$. If this holds for $\mathfrak{F} = \mathfrak{M}'$ we call B *functionally closed.*

5. 4. THEOREM. *Let $I = T^{-1}(0)$ and let $W(W_1)$ be the canonical homomorphism of B onto B/I (B/I onto $B/kh(I)$). Let T_1 be the algebraic isomorphism given by $T = T_1W$. Suppose that*

(i) *B is regular and functionally closed with respect to $h(I)$, and either*
(ii) *$T(R)$ is closed in $T(B)$ or*
(ii)′ *I is closed and $T(kh(I))$ is closed in $T(B)$.*

[7] The notation $u \circ v = u + v - uv$ is employed.

Then

(a) $W_1T_1^{-1}$ *is continuous,*[8]

(b) $\overline{T(B)} = T(B) + R_1$ *where* R_1 *is the radical of* $\overline{T(B)}$.

Assume (i) and (ii). By (ii), $R + I = T^{-1}[\overline{T(R)}]$ so that, by Theorem 5.3, $T^*(\mathfrak{R}') = h(I)$. Let γ be the mapping $x \to x(M' \mid h(I))$ of B into $C(h(I))$ and let $\pi: y \to y(N')$ be the Gelfand representation of B_1. By formula (1) the correspondence $V: V\gamma(x) = \pi(x)$ is a norm-preserving isomorphism of $\gamma(B)$ onto $\pi T(B)$. By (i) $\gamma(B)$ is a Banach algebra. Then $\pi T(B)$ is closed in $\pi(B_1)$. Now there is no loss of generality in taking $B_1 = \overline{T(B)}$. This choice gives $\overline{T(B)} \supset \pi^{-1}\pi T(B) \supset T(B)$. Inasmuch as π is continuous it follows that $B_1 = \pi^{-1}\pi T(B)$ from which (b) is evident.

The kernel of γ is $kh(I)$. Hence, ([2], p. 41), the Banach algebra $B/kh(I)$ is equivalent to $\gamma(B)$. Thus there is a number $c > 0$ such that

$$\| T(x) \| \geqq \sup_{N' \in \mathfrak{R}'} | T(x)(N') | = \| \gamma(x) \| \geqq c \| x + kh(I) \|$$

for all $x \in B$. Since $W_1T^{-1}(T(x)) = x + kh(I)$, (a) follows.

Now assume (i) and (ii)′. In view of Lemma 5.1, B/I is a regular Banach algebra. As a purely algebraic matter, ([7], p. 63), only maximal regular ideals in B which contain I do have images under W contained in maximal regular ideals of B/I. Let $\mathfrak{R}_1$ be the compactified space of maximal regular ideals of B/I. Then $W^*(\mathfrak{R}_1') = h(I)$. By (i) and formula (1) B/I is functionally closed. The radical of B/I is $Wkh(I)$, ([7], p. 63), and by (ii)′, $T_1(WkhI)$ is closed in $T(B) = T_1(B/I)$.

Therefore with B/I, T_1 and 0 (the zero of B/I) playing the roles of B, T, and I respectively (i) and (ii) are fulfilled. The above immediately gives (b) and the continuity of $W_2T_1^{-1}$ where W_2 is the canonical homomorphism of B/I onto $(B/I)(kh(0)) = (B/I)/(Wkh(I))$. The latter is algebraically isomorphic to $B/kh(I)$ by standard algebraic arguments where, moreover, the isomorphism preserves the quotient algebra norms. This yields (a).

5.5. COROLLARY. *In the notation of Theorem* 5.4, *suppose that* (i) *holds and* (ii)″ $I = kh(I)$. *Then* $\overline{T(B)} = T(B) \oplus R_1$ *and* T_1^{-1} *is continuous.*

From (ii)″ we obtain (ii)′ immediately. Also W_1 is in this case the identity. To complete the proof we need only show that $T(B) \cap R_1 = 0$.

[8] $B/kh(I)$ is given the usual quotient algebra norm which is possible since $kh(I)$ is closed in B.

As a purely algebraic matter (ii)″ gives semi-simplicity for B/I. Hence by Theorem 5. 3 and Lemmas 5. 1, 5. 2, $0 = T_1^{-1}(R_1)$. Since $T(B) = T_1(B/I)$ the proof is complete.

If B is a B^*-algebra and $T^{-1}(0)$ is closed, then well-known properties of B give the hypotheses of Corollary 5. 5. The conclusion in this case is given in [6]. So far as we know the conclusion that T_1^{-1} is continuous is an open question if B is a non-commutative B^*-algebra.

UNIVERSITY OF OREGON.

BIBLIOGRAPHY.

[1] R. Arens, "Approximation in, and representation of, certain Banach algebras," *American Journal of Mathematics*, vol. 71 (1949), pp. 763-790.

[2] S. Banach, *Theorie des operations linéaires*, Monografje Matematyzme, Warsaw (1932).

[3] I. Gelfand and M. Neumark, "On the imbedding of normed rings into the ring of operators in Hilbert space," *Receuil Mathématique* (*Matematičeskii Sbornik*) n. s., vol. 12 (1943), pp. 197-213.

[4] I. Gelfand and G. Šilov, "Uber verschiedene Methoden der Einführung der Topologie in die Menge der maximalen Idealen eines normierten Ringes," *ibid.*, vol. 9 (1941), pp. 25-39.

[5] E. Hille, *Functional analysis and semi-groups*, American Mathematical Society Colloquium Publications, vol. 31, New York (1948).

[6] I. Kaplansky, "Normed algebras," *Duke Mathematical Journal*, vol. 16 (1949), pp. 399-418.

[7] L. H. Loomis, *An introduction to abstract harmonic analysis*, D. Van Nostrand Co., New York, 1953.

[8] M. Neumark, "Rings with involution," *Uspehi Matematičeskii Nauk*, vol. 3 (1948), pp. 52-145 (Russian).

[9] ———, "Rings with involution," *American Mathematical Society Translation*, no. 25, New York, 1950.

[10] J. D. Newburgh, "The variation of spectra," *Duke Mathematical Journal*, vol. 18 (1951), pp. 165-176.

[11] C. E. Rickart, "The singular elements of a Banach algebra," *ibid.*, vol. 14 (1947), pp. 1063-1077.

[12] ———, "The uniqueness of norm problem for Banach algebras," *Annals of Mathematics*, vol. 51 (1950), pp. 615-628.

[13] ———, "On spectral permanence for certain Banach algebras," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 191-196.

[14] G. Šilov, "On regular normed rings," *Traveaux de l'Institut Mathématique Stekloff*, vol. 21 (1947). (Russian, English summary.)

LINEAR FUNCTIONALS AND CONTENT.*

By L. H. Loomis.

1. Introduction. The Riesz representation theorem for linear functionals, in a general form, may be stated as follows: To every non-negative linear functional I on the algebra $\boldsymbol{A}$ of all continuous real-valued functions on a compact Hausdorff space S, there corresponds a unique finite Baire (or Radon) measure μ such that $I(f) = \int f d\mu$ for every $f \in \boldsymbol{A}$.

If $\boldsymbol{A}$ and I are as above but S is replaced by a dense subset M, there is still a uniquely determined measure μ (on M) with respect to which I is integration, but now μ is not in general countably additive, but only finitely additive. Such a relationship was studied by Markoff [5], who took $\boldsymbol{A}$ to be the algebra of all bounded continuous functions on a normal space M. However, whenever I is defined like this on a uniformly closed algebra of bounded functions, the underlying space M can always be compactified and the existence of μ can be inferred from the Riesz theorem. The situation is different if I is defined only on a proper subset of $\boldsymbol{A}$. Compactification procedures—in fact, the use of toplogical arguments of any kind—now appear less natural than certain direct methods, and it seems best to drop all assumptions pertaining to a topology on S and continuity of the functions of $\boldsymbol{A}$. An early example of such a study is the well known paper of Daniell [2], which treats a non-negative linear functional whose domain is a vector lattice $\boldsymbol{L}$ of bounded functions on a set S, with no topological assumptions whatever. Daniell was interested in the extension of I to a functional behaving like a Lebesgue integral, and added a convergence postulate in order to get the Lebesgue convergence theorems. He did not try to find an associated measure on S. However, in a similar situation Stone [6] showed that the additional postulate that $f \cap 1 \in \boldsymbol{L}$ whenever $f \in \boldsymbol{L}$ is sufficient to imply the existence of a unique completely additive measure on S with respect to which I is integration. In this context, the role of the Daniell convergence axiom is to prove that the finitely additive μ is actually countably additive, and hence that the Riemann type integral $I(f) = \int f d\mu$ is actually a Lebesgue integral. As far as concerns the existence of μ and the relationship $I(f) = \int f d\mu$, it would seem best to leave the convergence axiom out, as did

* Received April 9, 1953.

Markoff in the paper cited above. This is in keeping with a recent tendency to study the relationship $I(f) = \int f d\mu$ in very general settings (see [1, 4]). The present paper is written in this spirit, but from a point of view which does not seem to have been adopted before.

We begin our considerations with a linear space $\boldsymbol{L}$ of real-valued functions on a set S and a non-negative linear functional I on $\boldsymbol{L}$. If the functions of $\boldsymbol{L}$ are bounded and include the constant functions, then it is reasonable to call I an integral even in this general situation. For, as is well known, a modified Hahn-Banach theorem enables us to extend I in many ways to a non-negative linear functional on the space of *all* bounded functions on S. If m is the finitely additive measure defined on the subsets of S by $m(A) = I(\chi_A)$, where χ_A is the characteristic function of A, then it is easy to see that $I(f)$ is in fact the integral $\int f dm$.

Such measures are of no particular interest in their relationship to the original functional I. There are many of them, and no one can be singled out as having an especially intimate connection with I. However, there may exist sets $A \subset S$ on which all of the measures m agree. That is, there may exist sets A whose measures are uniquely determined by the original functional I. The investigation of this uniquely determined measure μ and its relation to I is the program of the present paper, the main problem being the determination of conditions under which μ is adequate for the integration of the functions of $\boldsymbol{L}$.

An examination of the functional extension process mentioned above shows that the measure of A is uniquely determined by I if and only if for every positive ϵ there exist functions g and $h \in \boldsymbol{L}$ such that $g \leqq \chi_A \leqq h$ and $I(h) - I(g) \leqq \epsilon$, in which case $\mu(A)$ must be the common extreme l. u. b. $I(g) =$ g. l. b. $I(h)$. Such a set will be said to have *content*. (It is difficult to resist the temptation to call A a contented set.)

This consideration suggests a preliminary extension of the domain of I by the classical upper and lower integrals. For this extension it is not necessary that the functions of $\boldsymbol{L}$ be bounded or that $1 \in \boldsymbol{L}$. If an arbitrary function f is bounded above by a function h in $\boldsymbol{L}$, we define the upper integral $\bar{I}(f)$ as the greatest lower bound of $I(h)$ for all such h. Similarly, if there exists $g \in \boldsymbol{L}$ such that $g \leqq f$, the lower integral $\underline{I}(f)$ is defined as l. u. b. $I(g)$ for all such g. The difference $\bar{I}(f) - \underline{I}(f)$ will be called the *deficiency* of f, and denoted def (f). A function f with deficiency zero has only one possible value for its integral, $I(f) = \bar{I}(f) = \underline{I}(f)$, and if the domain of I is extended to the set $\boldsymbol{R}$ of all such f, then it is clear that $\boldsymbol{R}$ is a vector space and that the extended I is non-negative on $\boldsymbol{R}$. We shall call

$\boldsymbol{R}$ the *two-sided completion* of $\boldsymbol{L}$ (with respect to I), and we observe that the two-sided completion of $\boldsymbol{R}$ is again $\boldsymbol{R}$; that is, $\boldsymbol{R}$ is two-sidedly complete. The definition of content can now be rephrased: A set A has content if and only if $\chi_A \in \boldsymbol{R}$, in which case $\mu(A)$ is defined as $I(\chi_A)$.

Let $\boldsymbol{E}$ be the family of linear combinations of characteristic functions of sets having content. A function f will be said to be *Riemann summable* if it belongs to the two-sided completion of $\boldsymbol{E}$. That is, f is Riemann summable if, given ϵ, there exist step functions g and h in $\boldsymbol{E}$ such that $g \leqq f \leqq h$ and $I(h) - I(g) < \epsilon$. It follows that such a function f is bounded and belongs to $\boldsymbol{R}$, and that $I(f)$ is at once the least upper bound of all such numbers $I(g)$ and the greatest lower bound of all such numbers $I(h)$. Since g is of the form $g = \sum c_i \chi_{A_i}$, where the sets A_i have content, and therefore $I(g) = \sum c_i I(\chi_{A_i}) = \sum c_i \mu(A_i) = \int g d\mu$, we have $I(f) = \int f d\mu$ for a Riemann summable function in exactly the sense of the classical Riemann integral.

Supposing that the functions of $\boldsymbol{L}$ (and hence $\boldsymbol{R}$) are bounded, the optimum situation would be that in which the family $\boldsymbol{F}$ of sets having content is a Boolean algebra, or at least a Boolean ring, and is large enough so that all the functions of $\boldsymbol{L}$ (and hence $\boldsymbol{R}$) are Riemann summable.

Now it is clear that if $\boldsymbol{F}$ is a Boolean ring then $\boldsymbol{E}$ is a vector lattice, which in turn easily implies that if f and g are Riemann summable, then so are $f \cup g$ and $f \cap g$, where $f \cup g = \max(f, g)$ and $f \cap g = \min(f, g)$. Thus if $\boldsymbol{F}$ is to be a Boolean ring and $\boldsymbol{R}$ is to be precisely the set of Riemann summable functions then $\boldsymbol{R}$ must be a vector lattice. Another obviously necessary condition is the property used by Stone, namely, that if $f \in \boldsymbol{R}$ then $f \cap 1 \in \boldsymbol{R}$. These well known necessary conditions are not sufficient however, and we shall be largely concerned with the implications which they do have, and with the extra hypotheses which we must add to ensure the summability of the functions of $\boldsymbol{R}$.

2. Measurability and Riemann summability. We begin our systematic discussion with some definitions. If $f \geqq 0$ the *spectral family* of f is the one parameter family of sets F_λ of the form $F_\lambda = \{p : f(p) > \lambda\}$, $\lambda > 0$. Alternatively, we could have used "$\geqq$" in this definition. However, the other two one parameter families, having "$<$" and "$\leqq$" in their definitions instead of "$>$," and consisting therefore of the sets complementary to those in the above two families, would not in general do for our purpose because these complementary sets may be "too large" to have content.

A non-negative function f is *measurable* if all of its spectral sets F_λ

have content, except perhaps for a countable number of exceptional values of λ. The allowance for this exceptional set seems to be the correct modification of the usual definition of measurability when the measure in question is not completely additive. A function having both signs is said to be measurable if its positive and negative parts are measurable.

LEMMA 1. *Let $\boldsymbol{L}$ be a vector space of real-valued functions on a set S and let I be a non-negative linear functional on $\boldsymbol{L}$. If $f \cap 1 \in \boldsymbol{L}$ whenever $f \in \boldsymbol{L}$ then every $f \in \boldsymbol{L}$ is measurable.*

Proof. Let $0 < \lambda_0 < \lambda_1 < \cdots < \lambda_{m+1} < \infty$ be any finite sequence of positive numbers, and let g_n be the function $[(f \cap \lambda_n) - (f \cap \lambda_{n-1})]/(\lambda_n - \lambda_{n-1})$. Then $0 \leqq g_n \leqq 1$, $g_n = 0$ on the set where $f \leqq \lambda_{n-1}$ and $g_n = 1$ on the set F_{λ_n} where $f > \lambda_n$. Thus g_n is simultaneously a lower approximant for $\phi_{\lambda_{n-1}}$ (where ϕ_λ is the characteristic function of F_λ) and an upper approximant for ϕ_{λ_n}. Thus $\operatorname{def}(\phi_{\lambda_n}) \leqq I(g_n) - I(g_{n+1})$ and $\sum_1^m \operatorname{def}(\phi_{\lambda_n}) \leqq I(g_1)$. We now fix $\lambda_1 = a$, and, given ϵ, we choose the remaining λ_j from among the points λ at which the functions ϕ_λ have deficiencies greater than ϵ. Then $m\epsilon \leqq I(g_1)$ as above; that is, the number of points beyond a at which ϕ_λ has deficiency $\geqq \epsilon$ is bounded above by $I(g_1)/\epsilon$. Varying a and ϵ we see that the set of positive λ at which the deficiency of ϕ_λ is positive is at most countable. The sets F_λ for $\lambda < 0$ are accounted for by applying the above result to $-f$.

Remark. This proof holds as well for the other decreasing spectral family, composed of the sets $B_\lambda = \{p : f(p) \geqq \lambda\}$ when $\lambda > 0$.

Remark. Actually, it is only the two-sided completion $\boldsymbol{R}$ which we need to be closed under intersection with 1. While this property for $\boldsymbol{R}$ will most easily follow from the same property for $\boldsymbol{L}$, we shall later on consider one situation in which it evolves differently.

The hypothesis that $f \cap 1 \in \boldsymbol{L}$ for every $f \in \boldsymbol{L}$ almost implies by itself that $\boldsymbol{L}$ is a vector lattice. In fact, if $1 \in \boldsymbol{L}$ then the two conditions are equivalent, as can be seen from the relations $f \cap 0 = ((f+1) \cap 1) - 1$ and $f \cap g = g + ((f-g) \cap 0)$. If 1 is not in $\boldsymbol{L}$, but $f \cap 1$ is always in $\boldsymbol{L}$, then $\boldsymbol{F}$ is a Boolean ring and $\boldsymbol{E}$ is a vector lattice, for the spectral sets of $\chi_A + \chi_B$ include $A \cap B$ uncountably often, so that, by Lemma 1, $A \cap B$ must have content if A and B have content. Moreover, if in this case $\boldsymbol{L}$ satisfies any of our later conditions for improper Riemann summability, then it requires very little more to imply that $\boldsymbol{R}$ itself will be a vector lattice. For instance, suppose that $\lim_{\epsilon \to 0} I(g \cap \epsilon) \leqq 0$ for every $g \in \boldsymbol{L}$ and suppose also that every

$f \in \boldsymbol{L}$ is dominated by a non-negative $m \in \boldsymbol{L}$. Then $g \leqq f \cup 0 \leqq h$, where $g = f - f \cap \epsilon$ and $h = (f - f \cap \epsilon) + m \cap \epsilon$. Given δ, we choose ϵ so that $I(m \cap \epsilon) < \delta$, and then have $I(h) - I(g) = I(m \cap \epsilon) < \delta$, proving that $f \cup 0 \in \boldsymbol{R}$, and hence that $\boldsymbol{R}$ is a lattice.

If $f \cap 1$ is always in $\boldsymbol{L}$ our discussion to date gives us a finitely additive measure μ defined on quite a large Boolean ring of sets, but it does not yet follow that μ has any useful relationship to I, as the following examples show (see [4]).

Example 1. $\boldsymbol{L}$ is the vector lattice of continuous functions f on $[0, 1]$ such that $f(0) = 0$ and $f'(0)$ exists, and $I(f) = f'(0)$. I is then a non-negative linear functional. On the other hand, a subset A of $[0, 1]$ has content if and only if it is bounded away from 0, its content being always 0.

Example 2. $\boldsymbol{L}$ is the vector lattice of continuous functions on $[0, \infty)$ having derivatives at ∞ $(f'(\infty) = \lim_{x \to \infty} f(x)/x)$, and $I(f) = f'(\infty)$. Here a set A has content if and only if it is bounded, its content being always zero.

The failure of I to be reasonably related to μ in these two cases has different causes. In the second case it is due to the fact that the functions of $\boldsymbol{L}$ are in general unbounded, while in the first case the functions of $\boldsymbol{L}$ are all bounded but the constant functions are missing. However, if neither of these defects is present we shall see that every $f \in \boldsymbol{R}$ is Riemann summable.

We notice first that the presence of 1 in $\boldsymbol{L}$ (or in $\boldsymbol{R}$) has the very strong implication that I is a bounded linear functional in the uniform norm, with norm $I(1)$ $(I(f) \leqq I(\| f \|_\infty) = \| f \|_\infty I(1))$, and hence that $\boldsymbol{R}$ is uniformly closed. Also, $\boldsymbol{F}$ is now closed under complementation, for if A' is the complement of A then $\chi_{A'} = 1 - \chi_A$. The assumption that $f \geqq 0$ can now be dropped in the definition of the spectral family: the set where $f > -a$ is the complement of the set where $-f \geqq a$, and hence has content for all except a countable number of values of a.

LEMMA 2. *Let $\boldsymbol{L}$ be a vector space of real-valued functions on a set S, and let I be a non-negative linear functional on $\boldsymbol{L}$. If $1 \in \boldsymbol{L}$ then every bounded measurable function f is Riemann summable. More generally, if f is a bounded measurable function such that the set where $f \neq 0$ is included in a set A having content, then f is Riemann summable.*

Proof. Adding a suitable scalar multiple of χ_A to f, we can suppose that $f \geqq 0$ and that $F_\lambda = A$ for some positive λ. Taking $b \geqq$ l. u. b. f, let $0 < \lambda_0 < \lambda_1 < \cdots < \lambda_n = b$ be a subdivision of $[0, b]$ of norm $\leqq \epsilon$ such

that F_λ has content for every i and $F_{\lambda_0} = A$. Setting $g_i = \phi_{\lambda_{i-1}} - \phi_{\lambda_i}$, $g = \sum_1^n \lambda_{i-1} g_i$ and $h = \sum_1^n \lambda_i g_i$, we have $g, h \in \boldsymbol{E}$, $g \leqq f \leqq h$, and

$$I(h) - I(g) = \sum(\lambda_i - \lambda_{i-1}) I(g_i) \leqq \epsilon \sum I(g_i) = \epsilon\mu(A),$$

proving the lemma.

It is clear from the above proof that the hypothesis of measurability can be weakened to the assumption that the spectral sets F_λ have content for a dense set of λ.

COROLLARY 1. *If f is measurable, and $0 < a < b$ then $f \cap b - f \cap a$ is Riemann summable.*

Proof. The function $f \cap b - f \cap a$ is bounded by b and the set where it is different from zero is the spectral set F_a.

COROLLARY 2. *If $\boldsymbol{L}$ is closed under intersection with 1 and f is measurable and non-negative, then $f \cap g$ is Riemann summable for every Riemann summable g.*

Proof. The hypothesis that $f \cap 1 \in \boldsymbol{L}$ implies that $\boldsymbol{F}$ is a Boolean ring, as we saw in the discussion following Lemma 1. The result then follows from Lemma 2.

Remark. The expression of $\int f d\mu$ as the limit of sums $\sum \lambda_i(\mu(F_{\lambda_i}) - \mu(F_{\lambda_{i+1}}))$ is equivalent to $\int f d\mu = -\int_a^b \lambda d\alpha(\lambda)$, where α is the monotone decreasing function defined for all positive λ (except the values of λ in the countable exceptional set) by $\alpha(\lambda) = \mu(F_\lambda)$.

As already noted, Lemmas 1 and 2 imply the following theorem.

THEOREM 1. *If $\boldsymbol{L}$ is a vector-lattice of real-valued functions which contains the constant functions, and if I is a non-negative linear functional on $\boldsymbol{L}$, then the subsets of S having content form a Boolean algebra on which the content μ is finitely additive, the class of functions which are Riemann summable with respect to μ is precisely the set of bounded functions in $\boldsymbol{R}$, and $I(f) = \int f d\mu$ in the Riemann sense, for every such f.*

COROLLARY. *In the same situation, if f is a bounded function of $\boldsymbol{R}$ and if $\theta(x)$ is a monotone continuous function of the real variable x defined on a closed interval containing the range of f, then $\theta(f) \in \boldsymbol{R}$.*

Proof. The spectral sets of $\theta(f)$ are the same as those of f (with a

monotone change of parameter) and hence all except a countable number of them have content by Lemma 1. Lemma 2 then implies that $\theta(f) \in \boldsymbol{R}$.

Remark. It follows from the corollary that if $f \in \boldsymbol{R}$ is bounded then $f^2 \in \boldsymbol{R}$. Hence if f and g are bounded functions of $\boldsymbol{R}$ then

$$fg = [(f+g)^2 - (f-g)^2]/4 \in \boldsymbol{R},$$

so that the bounded functions of $\boldsymbol{R}$ form a uniformly closed algebra.

It now follows that if $f_1, \cdots, f_n$ are bounded functions of $\boldsymbol{R}$, then so is $\alpha(f_1, \cdots, f_n)$, where α is any continuous function of n variables. For, by the Weierstrass theorem, $\alpha(f_1, \cdots, f_n)$ is a uniform limit of polynomials in $f_1, \cdots, f_n$.

3. Improper integration. We turn now to the problem of the integration of unbounded functions or of bounded functions when the constants are not present. As the two examples above suggest, the difficulties here are manifestations of two aspects of the same phenomenon: We cannot now obtain two-sided approximations by step functions, in the one case because the function in question is unbounded, and in the other case because the total content of S may be unbounded. Our theory therefore must of necessity be a theory of one-sided approximation, that is, of improper Riemann integration. As usual, we take as our point of departure a vector space $\boldsymbol{L}$ of real-valued functions which is closed under intersection with 1, and a non-negative linear functional I on $\boldsymbol{L}$.

We say that a non-negative function f is *improperly Riemann summable* if it is measurable and if there is a constant k such that $I(g) < k$ whenever g is a Riemann summable function such that $g \leqq f$. The improper Riemann integral $\int f d\mu$ is then defined as the least upper bound of such integrals $I(g)$ $(= \int g d\mu)$. If $\boldsymbol{L}^+$ is the set of non-negative functions of $\boldsymbol{L}$, then it follows from Lemma 1 and the non-negativity of I that any $f \in \boldsymbol{L}^+$ is improperly Riemann summable, with $\int f d\mu \leqq I(f)$, and one of our problems is the discovery of conditions which will imply the equality. We first set down a few rather obvious equivalent formulations of the definition of $\int f d\mu$.

LEMMA 3. *If f is non-negative and improperly summable then the function g occurring in the definition of $\int f d\mu$ can be restricted to functions of any one of the following forms:*

(1) *the step functions of $\boldsymbol{E}$ which are $\leqq f$;*

(2) *$f \cap b\chi_A$, where A has content and $b > 0$;*

(3) *$f \cap b\phi_a$, where ϕ_λ is the characteristic function of the spectral set F_λ.*

Proof. (1) is obvious. (2) follows from the fact that $f \cap b\chi_A$ is Riemann summable by Lemma 2, and that any Riemann summable function is bounded and is zero except on a set A having content. To prove (3) we choose any step function $h \in \boldsymbol{E}$ such that $0 \leqq h \leqq f$. If a is the minimum positive value of h and b is the maximum value of h, then $h \leqq b\phi_a$, and (3) follows from (1) and the Riemann summability of $f \cap b\phi_a$.

We obtain two other slightly stronger conditions from the following lemma.

LEMMA 4. *If f is non-negative and improperly summable, then $a\mu(F_a) \to 0$ as $a \to 0$ and as $a \to \infty$.*

Proof. Given ϵ, choose $h \in \boldsymbol{E}$ so that $0 \leqq h \leqq f$ and $\int f d\mu - I(h) < \epsilon/2$. If A is the set where $h \neq 0$ and a is a positive number such that $a\chi_A \leqq h$, then $A \subset F_a$. Since $a(\phi_a - \chi_A) + h \leqq f$, we have $a\mu(F_a - A) < \epsilon/2$, and it is only necessary to take a so that $a\mu(A) < \epsilon/2$ to get $a\mu(F_a) < \epsilon$.

The proof is similar as $b \to \infty$. If $b \geqq 2 \max h$, then $h + (b/2)\phi_b \leqq f$ and hence $(b/2)\mu(F_b) < \epsilon/2$.

COROLLARY 1. *In the definition of the improper Riemann integral of f, g may be restricted to functions of either of the following forms:*

(4) $f \cap b - f \cap a$, *where* $0 < a < b$;

(5) $f \cap b(\phi_a - \phi_b)$, *where* $0 < a < b$.

Proof. Since $f \cap b - f \cap a = (f \cap b\phi_a) - a\phi_a$ and $f \cap b(\phi_a - \phi_b) = (f \cap b\phi_a) - b\phi_b$, the corollary follows from the lemma and (3) of Lemma 3.

COROLLARY 2. *A non-negative measurable function f is improperly Riemann summable if and only if the integral $\int_0^\infty \lambda d\alpha(\lambda)$ is convergent in the usual sense, in which case $\int f d\mu = -\int_0^\infty \lambda d\alpha(\lambda)$.*

Proof. This follows from (5) above and the fact that $I(f \cap b(\phi_a - \phi_b)) = -\int_a^b \lambda d\alpha(\lambda)$, where $\alpha(\lambda) = \mu(F_\lambda)$.

Each of the conditions (1)-(5) translates into a necessary and sufficient condition for the equation $I(f) = \int f d\mu$ to hold, and others can easily be found. We collect a few in the following lemmas, where, for simplicity, the two kinds of improper behavior are separated.

LEMMA 5. *If f is a bounded non-negative improperly summable function then the equality $I(f) = \int f d\mu$ will hold:*

(a) *if and only if $I(f \cap a) \to 0$ as $a \to 0$;*

(b) *if and only if for every ϵ there exists $g \in \boldsymbol{L}^+$ and a set A having content such that $I(g) < \epsilon$ and $g \geqq f$ on A';*

(c) *if the set of numbers $\{I(nf \cap 1)\}$ is bounded above;*

(d) *if there exists $h \in \boldsymbol{L}^+$ such that $f/h \to 0$ as $f \to 0$.*

Proof. (a) is the direct translation of the condition (4) above, since $f \cap b = f$ for suitable b. Then we can take $g = f \cap a$ and $A = F_a$ to prove that (b) is necessary. Conversely, if (b) holds, and if $b = \text{l. u. b. } f(p)$, then $f \cap b\chi_A \geqq f - g$, so that the sufficiency of (b) follows from (2) of Lemma 3. The assumption in (c) implies that $I(f \cap 1/n) \to 0$, and (a) can be applied. Finally, if h exists as in (d), then given ϵ there exists δ such that $f < \epsilon h$ when $f \leqq \delta$. We can then take g in (b) as ϵh and A as F_δ.

LEMMA 6. *If $1 \in \boldsymbol{L}$ and f is a non-negative improperly summable function, then the equality $I(f) = \int f d\mu$ will hold:*

(a) *if and only if $I(f \cap b) \to I(f)$ as $b \to \infty$;*

(b) *if and only if for every ϵ there exists $g \in \boldsymbol{L}^+$ and b such that $I(g) < \epsilon$ and $g \geqq f$ on F_b;*

(c) *if there exists $h \in \boldsymbol{L}^+$ such that $f/h \to 0$ as $f \to \infty$.*

Proof. The proof is similar to that of the above lemma, and will be omitted.

The last conditions in the above two lemmas are in a sense necessary as well as sufficient, but this cannot be seen until another completion process for $\boldsymbol{L}$ has been developed, and the discussion will be postponed until Section 5.

4. Integrals on algebras of functions.

THEOREM 2. *If $\boldsymbol{A}$ is an algebra of bounded functions on a set S, if $1 \in \boldsymbol{A}$, and if I is a non-negative linear functional on $\boldsymbol{A}$, then the two-sided completion $\boldsymbol{R}$ is a lattice as well as an algebra and all the conclusions of Theorem 1 follow.*

Proof. This follows, as in the proof of the Stone-Weierstrass theorem, from the fact that $f \cup g = (f + g + |f - g|)/2$, that $y = |x|$ can be uniformly approximated by a polynomial in x^2 over any given finite interval,

and that because $1 \in \boldsymbol{A}$ the two-sided completion $\boldsymbol{R}$ includes the uniform closure of $\boldsymbol{A}$.

If the algebra $\boldsymbol{A}$ does not include the constant functions we have an improper integral situation as before. $\boldsymbol{R}$ cannot now be proved to be a lattice, but, if every $f \in \boldsymbol{A}$ is dominated by a function $g \in \boldsymbol{A}^+$, then we can show that $f \cap 1$ is always in $\boldsymbol{R}$.

LEMMA 7. *Let $\boldsymbol{A}$ be an algebra of bounded real-valued functions on a set S such that every function in $\boldsymbol{A}$ is a difference of non-negative functions of $\boldsymbol{A}$, and let I be a non-negative linear functional on $\boldsymbol{A}$. Then $f \cap 1$ belongs to the two-sidede completion $\boldsymbol{R}$ for every $f \in \boldsymbol{A}$.*

Proof. By hypothesis there exists $m \in \boldsymbol{A}$ such that $|f| \leqq m$. Let $[a, b]$ include the range of f, and let θ be the real-valued function on $[a, b]$ defined by $\theta(x) = 1$ if $x \leqq 1$ and $\theta(x) = 1/x$ if $x > 1$. Given ϵ, the classical Weierstrass theorem furnishes a polynomial P such that $|P(x) - \theta(x)| < \epsilon$ on $[a, b]$. Since $f\theta(f) = f \cap 1$, $fP(f) \in \boldsymbol{A}$ and $|f| \leqq m$, we have $g \leqq f \cap 1 \leqq h$, where $g = fP(f) - \epsilon m$, $h = fP(f) + \epsilon m$, and $I(h) - I(g) \leqq 2\epsilon I(m)$. Letting $\epsilon \to 0$, we see that $f \cap 1 \in \boldsymbol{R}$.

Thus every $f \in \boldsymbol{A}$ is measurable, and we can apply our results on improper summability. In particular, we have in more general form the main theorem of Arens [1].

THEOREM 3. *Let $\boldsymbol{A}$ be an algebra of bounded real-valued functions on a set S such that every $f \in \boldsymbol{A}$ is a difference of non-negative functions in $\boldsymbol{A}$. Let I be a non-negative linear functional on $\boldsymbol{A}$ and suppose that for every $f \in \boldsymbol{A}^+$, $I(f) = \text{l. u. b. } I(fu)$ where u ranges over the functions of $\boldsymbol{A}$ such that $0 \leqq u \leqq 1$. Then $I(f) = \int f d\mu$ for every $f \in \boldsymbol{A}$.*

Proof. Given $f \in \boldsymbol{A}^+$, we have to establish any one of the conditions of Lemma 5. But, given ϵ, we can choose $u \in \boldsymbol{A}$ such that $0 \leqq u \leqq 1$ and $I(f) - I(uf) < \epsilon/2$. Let A be the set on which $u > \delta$, δ being chosen so that A has content and so that $\delta I(f) < \epsilon/2$, and set $g = f(1 + \delta - u)$. Then $g \geqq f$ on A' and $I(g) = I(f) - I(fu) + \delta I(f) < \epsilon$, so that (b) of Lemma 5 holds, q. e. d.

The condition $I(f) = \text{l. u. b. } I(fu)$ of Theorem 3 has been studied in great detail in a different context by Dieudonné [3].

We can also consider algebras of unbounded functions, but, if we are to avoid the seemingly impossible difficulties inherent in such a generalization

of the determined moment problem, we shall have to assume that $\boldsymbol{A}$ is a lattice. The condition (c) of Lemma 6 is automatically met here, from the simple fact that $f \in \boldsymbol{A}$ implies $f^2 \in \boldsymbol{A}$, and, supposing for the moment that $1 \in \boldsymbol{A}$, it follows that $I(f) = \int f d\mu$ for every $f \in \boldsymbol{A}$. If the constants are not present, then one of the conditions of Lemma 5 must of course be assumed.

5. One-sided completion. We turn now to a deeper study of improper integrability, in which we consider the problem in the light of a second natural extension of I, this time by means of a one-sided completion process. The values of I for the new functions thus added are no longer uniquely determined in the sense of being forced by the old values, but they are uniquely determined in the sense of being explicitly and naturally defined.

We start with a vector-lattice $\boldsymbol{R}$ of real-valued functions which is two-sidedly complete with respect to a non-negative linear functional I. The non-negative part $\boldsymbol{U}^+$ of the extended class $\boldsymbol{U}$ is defined as follows: $f \in \boldsymbol{U}^+$ if and only if $f \geqq 0$, $f \cap g \in \boldsymbol{R}$ for every $g \in \boldsymbol{R}$, and l. u. b. $\{I(f \cap g) : g \in \boldsymbol{R}\} < \infty$. $I(f)$ is defined as the above least upper bound, and it is clear that $\boldsymbol{U}^+$ is a lattice and is closed under addition and multiplication by non-negative constants. Then $\boldsymbol{U}$ is the vector space of differences $f_1 - f_2$ of functions in $\boldsymbol{U}^+$ and $I(f_1 - f_2)$ is defined as $I(f_1) - I(f_2)$. Straightforward calculation shows I to be uniquely defined, linear and non-negative on $\boldsymbol{U}$.

Our first goal is to show that $\boldsymbol{U}$ is both one-sidedly and two-sidedly complete.

LEMMA 8. *If $f, g \in \boldsymbol{U}^+$ and $f \leqq g$ then $g - f \in \boldsymbol{U}^+$.*

Proof. We have to show that $(g - f) \cap h \in \boldsymbol{R}$ for every $h \in \boldsymbol{R}$. Given ϵ, choose $k \in \boldsymbol{R}$ such that $I(g) - I(g \cap k) < \epsilon$. Now $g \cap k - f \cap k \leqq g - f$. Also, the function $[g \cap (h + k)] - (f \cap k)$ is equal to $g - (f \cap k) \geqq g - f$ if $g \leqq h + k$, and to $h + k - (f \cap k) \geqq h$ if $g \geqq h + k$. Thus

$$h \cap (g \cap k - f \cap k) \leqq h \cap (g - f) \leqq h \cap ([g \cap (h + k)] - (f \cap k)),$$

where the extreme members belong to $\boldsymbol{R}$ (by the definition of $\boldsymbol{U}^+$) and the difference of the extremes is bounded by $g \cap (h + k) - g \cap k$. Since $I(g \cap (h + k)) - I(g \cap k) < \epsilon$ by the definition of k, we have proved that $h \cap (g - f)$ lies in the two-sided completion of $\boldsymbol{R}$, which is of course $\boldsymbol{R}$ itself. This is what we set out to prove.

THEOREM 4. *$\boldsymbol{U}$ is a vector lattice including $\boldsymbol{R}$ which is both two-sidedly and one-sidedely complete.*

Proof. We have already observed that $\boldsymbol{U}$ is a vector space including $\boldsymbol{R}$ and that I is a non-negative linear functional on $\boldsymbol{U}$. We continue with the fact that if $f \in \boldsymbol{U}$ then $f^+ = f \cup 0 \in \boldsymbol{U}^+$, for if $f = g - h$, where g and $h \in \boldsymbol{U}^+$, then $f^+ = g \cup h - h \in \boldsymbol{U}^+$ by the preceding lemma. Thus $\boldsymbol{U}^+$ is indeed the non-negative part of $\boldsymbol{U}$. Thus $|f| \in \boldsymbol{U}$ and $f \cup g = (f + g + |f - g|)/2 \in \boldsymbol{U}$.

Consider, now, any function f in the two-sided completion of $\boldsymbol{U}$; we may as well suppose f to be non-negative. Given ϵ, there exist by hypothesis functions $g, h \in \boldsymbol{U}^+$ such that $g \leqq f \leqq h$ and $I(h) - I(g) < \epsilon$. If $k \in \boldsymbol{R}$, it follows that $g \cap k \leqq f \cap k \leqq h \cap k$ and that $I(h \cap k) - I(g \cap k) < \epsilon$. Since $g \cap k$ and $h \cap k$ belong to $\boldsymbol{R}$, it follows that $f \cap k$ belongs to the two-sided completion of $\boldsymbol{R}$, which is $\boldsymbol{R}$ itself, and hence that f belongs to $\boldsymbol{U}^+$. Thus $\boldsymbol{U}$ is two-sidedly complete.

Finally, let f be any non-negative function in the one-sided completion of $\boldsymbol{U}$. That is, if $g \in \boldsymbol{U}^+$ then $f \cap g \in \boldsymbol{U}^+$ and l. u. b. $I(f \cap g) \leqq l < \infty$. But then, if $g \in \boldsymbol{R}$, $f \cap g = (f \cap g) \cap g = h \cap g$, where $h \in \boldsymbol{U}$, so that $f \cap g = h \cap g \in \boldsymbol{R}$ and hence $f \in \boldsymbol{U}$. Thus $\boldsymbol{U}$ is one-sidedly complete. This completes the proof of the theorem.

Since we have a larger two-sidedly complete space than before, we must consider the possibility of there being new sets having content. Suppose, then, that $\chi_A \in \boldsymbol{U}^+$. Since $\chi_A \cap g \in \boldsymbol{R}$ whenever $g \in \boldsymbol{R}$, and, in particular, whenever $g = \chi_B$ where $B \in \boldsymbol{F}$, we see that A has the property that $A \cap B \in \boldsymbol{F}$ whenever $B \in \boldsymbol{F}$. If there exists $B \in \boldsymbol{F}$ such that $A \subset B$, it follows that $A = A \cap B \in \boldsymbol{F}$ and A is not a new set. Thus the only new sets having content are sets A which are included in no sets of $\boldsymbol{F}$, but are such that $A \cap B \in \boldsymbol{F}$ for every $B \in \boldsymbol{F}$ and l. u. b. $\mu(A \cap B) < \infty$.

If $1 \in \boldsymbol{R}$, so that $S \in \boldsymbol{F}$, it follows that there can be no new sets having content.

If $1 \notin \boldsymbol{R}$, but $\boldsymbol{R}$ is closed under intersection with 1, it may happen that l. u. b. $\{I(f \cap 1) : f \in \boldsymbol{R}\} < \infty$, so that $1 \in \boldsymbol{U}^+$. This is clearly equivalent to I being bounded in the uniform norm, and implies that $\boldsymbol{R}$ and $\boldsymbol{U}$ are both uniformly closed. The converse also follows.

LEMMA 9. *If $\boldsymbol{R}$ is uniformly closed and closed under intersection with* 1, *then I is bounded and* $1 \in \boldsymbol{U}$.

Proof. If I is not bounded there exists a sequence A_n of disjoint sets having content such that $\sum \mu(A_n) = \infty$. Let c_n be a sequence of constants such that $c_n \downarrow 0$ and $\sum c_n \mu(A_n) = \infty$. The function $f = \sum c_n \chi_{A_n}$ is then in the uniform closure of $\boldsymbol{R}$, since it is approximated uniformly by its partial

sums, but cannot belong to $\boldsymbol{R}$ (or $\boldsymbol{U}$), since its integral is infinite. Therefore I is bounded.

We denote the enlarged $\boldsymbol{F}$ by $\boldsymbol{B}$, and suppose from now on that $\boldsymbol{R}$ is closed under intersection with 1, so that the functions of $\boldsymbol{R}$ and $\boldsymbol{U}$ are at least improperly summable (with respect to $\boldsymbol{F}$ and $\boldsymbol{B}$ respectively). We then have the theorem:

THEOREM 5. *If the equality* $\int f d\mu = I(f)$ *holds for every* $f \in \boldsymbol{R}$ *then it holds for every* $f \in \boldsymbol{U}$, *and* $\boldsymbol{U}$ *is precisely the set of improperly summable functions.*

Proof. The persistence of the identity $\int f d\mu = I(f)$ follows at once from the definitions of $\int f d\mu$ and $\boldsymbol{U}$, and we have only to prove that if f is a non-negative function which is improperly summable, then $f \in \boldsymbol{U}^+$. By definition, f is measurable and has the property that l. u. b. $\{I(g)\} = l < \infty$ for all Riemann summable functions $g \leqq f$. Let h be any function of $\boldsymbol{R}^+$, and, given ϵ, choose an $\boldsymbol{F}$-Riemann summable function $g \leqq h$ such that $I(h) - I(g) < \epsilon$. Since $f \cap h - f \cap g \leqq h - g$, we have $f \cap g \leqq f \cap h \leqq (h - g) + f \cap g$. Now $h - g \in \boldsymbol{R}$, and $f \cap g$ is an $\boldsymbol{F}$-Riemann summable function and so belongs to $\boldsymbol{R}$. The above inequality thus shows that $f \cap h$ belongs to the two-sided completion of $\boldsymbol{R}$, and hence to $\boldsymbol{R}$. The same inequality yields $I(f \cap h) \leqq I(h - g) + I(f \cap g) \leqq \epsilon + l$. Thus $f \cap h \in \boldsymbol{R}$ and $I(f \cap h) \leqq l$ for every $h \in \boldsymbol{R}$, so that $f \in \boldsymbol{U}^+$ by definition.

We can now prove, in terms of $\boldsymbol{U}$, that the sufficient condition of Lemma 5, (d), and Lemma 6, (c) is also necessary.

THEOREM 6. *Let* $\boldsymbol{U}$ *be a vector lattice of functions which is both two-sidedly and one-sidedly complete with respect to a non-negative linear functional* I, *and which is closed under intersection with* 1. *Then* $I(f) = \int f d\mu$ *for every* $f \in \boldsymbol{U}$ *if and only if for every* $f \in \boldsymbol{U}^+$ *there exists* $g \in \boldsymbol{U}^+$ *such that* $f/g \to 0$ *as* $f \to 0$ *or* ∞.

Proof. We have already seen that the existence of such a function g guarantees the identification of $\int f d\mu$ with $I(f)$. Conversely, if $I(f) = \int f d\mu$ for every $f \in \boldsymbol{U}$ then, by the above theorem, it is sufficient to exhibit a g with the desired properties which is improperly summable. Now $I(f) = -\int_0^\infty \lambda d\alpha(\lambda)$, where $\alpha(\lambda) = \mu(F_\lambda)$. Let $\theta(\lambda)$ be any strictly increasing continuous functions such that $\lambda/\theta(\lambda) \to 0$ as $\lambda \to 0$ or ∞ and

such that $\int_0^\infty \theta(\lambda)\,d\alpha(\lambda)$ converges. The function $g = \theta(f)$ is measurable, since it has the same spectral sets as f. If $\{G_\zeta\}$ is its spectral family and $\beta(\zeta) = \mu(G_\zeta)$, then $\int g\,d\mu = -\int_0^\infty \zeta\,d\beta(\zeta) = -\int_0^\infty \theta(\lambda)\,d\alpha(\lambda) < \infty$, so that g is improperly summable by Corollary 2 of Lemma 4. Finally, the definition of θ ensures that $f/g \to 0$ as $f \to 0$ or ∞, and this finishes the proof of the theorem.

6. We conclude with some remarks about the role of the assumption that $f \cap 1 \in \boldsymbol{L}$ for all $f \in \boldsymbol{L}$ when $\boldsymbol{L}$ is already assumed to be a vector lattice. In a sense, this assumption excludes irregularities which can occur when there aren't enough points in S. For, supposing that the functions of $\boldsymbol{L}$ are all bounded, standard weak compactification arguments allow us to enlarge S and to extend the functions of $\boldsymbol{L}$ uniquely so as to become a vector lattice of continuous functions on a compact Hausdorff space. With S thus enlarged (to $\bar{S}$, say) the hypothesis that $f \cap 1 \in \boldsymbol{L}$ can be greatly weakened.

First of all we may as well assume that $\boldsymbol{L}$ separates the points of $\bar{S}$, since unseparated points can always be identified. Secondly, we must avoid the situation where $\boldsymbol{L}$ is a one dimensional space consisting of the scalar multiples of a single function; it is obvious that no notion of content is available here. Nor can we allow this situation to exist on any subset of $\bar{S}$. That is, distinct points must be treated differently by different functions of $\boldsymbol{L}$; more exactly, if $p \neq q$ then there must exist $f, h \in \boldsymbol{L}$ such that $f(p)h(q) \neq f(q)h(p)$.

The interesting fact is that the above weak necessary condition actually implies that $f \cap 1$ belongs to the two-sided completion of $\boldsymbol{L}$ for every $f \in \boldsymbol{L}$. We can see this as follows. Given a non-negative function $f \in \boldsymbol{L}$, let $\boldsymbol{L}(f)$ be the sublattice of functions of $\boldsymbol{L}$ which are dominated by multiples of f. The quotients g/f, $g \in \boldsymbol{L}(f)$, form a vector lattice $\boldsymbol{L}'$ of bounded functions on the locally compact space S_0 obtained by excluding from $\bar{S}$ the closed set where $f = 0$, and $1 \in \boldsymbol{L}'$. Moreover our separation hypothesis above implies that $\boldsymbol{L}'$ separates points in S_0. We define J on $\boldsymbol{L}'$ by $J(g/f) = I(g)$; J is clearly a non-negative linear functional. Recompactifying, we can suppose that $\boldsymbol{L}'$ is a vector lattice of bounded continuous functions on a compact set $\bar{S}_0$, separating the points of this space and containing the constant functions. The two-sided extension with respect to J includes the uniform closure of $\boldsymbol{L}'$, and the Stone-Weierstrass theorem implies that the uniform closure of $\boldsymbol{L}'$ is the algebra of all continuous functions on $\bar{S}_0$. Going back, $(f \cap 1)/f$ is a con-

tinuous function on S_0 which is identically 1 in a neighborhood of ∞, and therefore remains continuous on the compactification $\bar{S}_0$. It therefore belongs to the two-sided extension with respect to J. That is, $f \cap 1$ belongs to the two-sided extension with respect to I, q. e. d.

HARVARD UNIVERSITY.

BIBLIOGRAPHY.

See [4] below for further bibliography.

[1] R. Arens, "The representation of functionals by integrals," *Duke Mathematical Journal*, vol. 17 (1950), pp. 499-506.

[2] P. J. Daniell, "A general form of integral," *Annals of Mathematics*, vol. 19 (1917), pp. 279-294.

[3] J. Dieudonné, "Sur le théorème de Lebesgue-Nikodyn, II," *Bulletin de la Société Mathématique de France*, vol. 72 (1944), pp. 193-239.

[4] E. Hewitt, "Integral representation of certain linear functionals," *Arkiv for Matematik*, vol. 2 (1951), pp. 269-282.

[5] A. Markoff, "On mean values and exterior densities," *Matematiceskii Sbornik*, vol. 46 (1938), pp. 165-191.

[6] M. H. Stone, "Notes on integration, II," *Proceedings of the National Academy of Sciences*, U. S. A., vol. 34 (1948), pp. 447-455.

ON A THEOREM OF BÔCHER IN THE THEORY OF ORDINARY LINEAR DIFFERENTIAL EQUATIONS.*

By AUREL WINTNER.

Let A be an n by n matrix, defined as a continuous function $A = A(t)$ for large positive t, and consider the linear differential equation

$$x' = A(t)x \tag{1}$$

for the n-rowed vector x. If there belongs to every constant vector c a solution vector $x(t)$ of (1) satisfying $x(t) \to c$ as $t \to \infty$ (that is, if the "initial condition" $x(\infty)$ can be assigned arbitrarily), then $A(t)$ or (1) will be said to be of *type* (*) or to have (*)-*character*. It is clear from the principle of superposition (and from the existence of n linearly independent solution vectors) that, if $A(t)$ is of type (*), then there belongs to every c exactly one $x(t)$, and $x(\infty) = \lim x(t)$ must exist for every solution $x(t)$.

For the same reasons, $A(t)$ is of type (*) if and only if there exists a (unique) fundamental matrix $X(t)$ satisfying

$$X(t) \to E \text{ as } t \to \infty, \tag{2}$$

where E is the n-rowed unit matrix. In fact, a fundamental matrix X is defined as any n by n matrix the columns of which are linearly independent solution vectors x of (1); so that

$$X' = A(t)X \tag{3}$$

and $\det X(t) \neq 0$. The latter inequality holds for every $t < \infty$ if it holds for a single t; in which case it holds for $t = \infty$ also, provided that $A(t)$ is of type (*). It is also clear that if $A(t)$ is said to possess (**)-character when the limit $X(\infty)$ exists for some (hence for every) fundamental matrix $X(t)$, then the (**)-character of $A(t)$ and the additional condition $\det X(\infty) \neq 0$ (for some, hence for every, fundamental matrix) together are necessary and sufficient for the (*)-character of $A(t)$.

---

* Received June 20, 1953.

If $f(t)$, where const. $\leqq t < \infty$, is a continuous (scalar, vector or matrix) function for which the improper integral

$$\int^{\infty} f(t)\,dt \text{ is convergent, i. e., } \lim_{T\to\infty} \int^{T} f(t)\,dt \text{ exists}$$

(as a finite limit), then $f(t)$ will be called of class R. Thus $f(t) \varepsilon R$ is necessary but not sufficient for $f(t) \varepsilon L$, if L denotes the case $p = 1$ of the class L^p, that is, of the class consisting of those (continuous) matrix functions $f(t)$ all elements of which have *absolutely* integrable p-th powers on the half-line const. $\leqq t < \infty$.

Suppose that

$$A(t) \varepsilon R \tag{4}$$

holds for the coefficient matrix of (1) (i. e., that the matrix function

$$A^{\circ}(t) = \int_{t}^{\infty} A(s)\,ds, \text{ where } \int_{t}^{\infty} = \lim_{T\to\infty} \int_{t}^{T}, \tag{5}$$

exists, hence

$$A^{\circ}(t) = o(1) \tag{6}$$

as $t \to \infty$). Then, in order that (1) has (*)-character, any of the following three conditions is sufficient:

$$A(t) \varepsilon L \tag{I}$$

(a condition under which the additional requirement (4)-(5) is superfluous);

$$A(t) \varepsilon L^p \text{ and } A^{\circ}(t) \varepsilon L^q \tag{II}$$

for some $p > 1$ and the corresponding $q = p/(p-1) > 1$; finally,

$$A(t) = O(1) \text{ and } A^{\circ}(t) \varepsilon L, \tag{III}$$

where the O-symbol refers to $t \to \infty$.

The (*)-criterion (I) goes back to Bôcher, appearing in the literature for over fifty years (for references, cf. [1], p. 486, footnotes 57 and 58). Since (I) will be used below, a simple proof of its sufficiency for the (*)-character of $A(t)$ will first be given.

It is easy to prove (see [3]) that there belongs to every solution vector $x(t) \not\equiv 0$ of (1) a positive number ξ satisfying $|x(t)| \to \xi$, as $t \to \infty$, whenever $A(t)$ satisfies the pair of conditions $\lambda(t) \varepsilon R$, $\mu(t) \varepsilon R$, where λ denotes the least, and μ the greatest, eigenvalue of the Hermitian matrix $\frac{1}{2}(A + A^*)$.

On the other hand, it is clear that $A(t) \varepsilon L$ is sufficient for $\lambda(t) \varepsilon L$ $(\supset R)$ and $\mu(t) \varepsilon L$ $(\supset R)$. Hence $A(t) \varepsilon L$ implies that, if $x(t)$ is any solution vector, then $x(t) = O(1)$, hence $A(t)x(t) \varepsilon L$. In view of (1), this means that $x'(t) \varepsilon L$. In particular $x'(t) \varepsilon R$, i. e., the limit $x(\infty)$ exists. Finally, $x(\infty) \neq 0$ unless $x(t) \equiv 0$, since ξ is 0 only in the latter case. This proves (I).

(II) and (III) were proved in [2].

It is clear that none of the three assumptions (I), (II), (III) on $A(t)$ is contained in the other two, and that the two conditions of (II) belonging to distinct values of p are all distinct; so that all these sufficient criteria for the (*)-character of (1) are independent. It will however be shown that, by an adaptation of the formal device by means of which the function (5) was introduced in [2], it is possible to deduce from the first of the criteria (I)-(III) *one* criterion which contains all three (and more). The criterion in question is condition (i) in the following assertion:

In order that (1) *be of type* (*), *it is sufficient that*

(i) $$A^\circ(t)A(t) \varepsilon L$$

(where (4), i. e., the existence of the function (5), is part of the assumption).

Since (I) is sufficient for (4)-(6), it is sufficient for (i). Similarly, (i) follows from (III) also. Finally, (i) follows from (II) by Hölder's inequality. Thus the (*)-criterion (i) contains all of the preceding ones.

Since the factors of the matrix product occurring in (i) are not commutable (in general), it is by no means clear that (i) *can be replaced by*

(ii) $$A(t)A^\circ(t) \varepsilon L$$

as a sufficient condition for the (*)*-character of* (1). That (ii) proves to be sufficient, will be due to the fact that if $X_1 = X_1(t)$ is a fundamental matrix belonging to a coefficient matrix $A_1(t)$, then the inverse X_1^{-1} of X_1 is a fundamental matrix $X_2 = X_2(t)$ belonging to the coefficient matrix $A_2(t)$, where $A_2 = -X_1^{-1}A_1X_1$. This fact can be verified from the rule

(7) $$(F^{-1})' = -F^{-1}F'F^{-1},$$

which is valid for every differentiable matrix function $F = F(t)$ of non-vanishing determinant.

The proof of the sufficiency of (i) proceeds as follows: Since (5) implies that $A = (-A^\circ)'$, the equation (3) can be written in the form

(8) $$X' = (-A^\circ)'X \equiv -(A^\circ X)' + A^\circ X' = -(A^\circ X)' + A^\circ AX,$$

by (3). Since (8) can be written in the form $Y' = A^\circ AX$, where

$$Y = (E + A^\circ)X, \tag{9}$$

it follows that $Y' = CY$, where

$$C = A^\circ A(E + A^\circ)^{-1}. \tag{10}$$

Accordingly, (3) is equivalent to

$$Y' = C(t)Y \tag{11}$$

by virtue of the definitions (9)-(10), provided that t is large enough. The latter proviso is needed in order that (6) assures the existence of the reciprocal which is the third factor in (10).

Suppose that $A(t)$ satisfies condition (i). Then (10) and (6) imply that $C(t) \in L$. Hence, if the (*)-criterion (I) for (3) is applied to (11), it follows that $C(t)$ is of type (*). Corresponding to the formulation (2) of the (*)-property of (3), this means that (11) has a solution $Y = Y(t)$ satisfying $Y(t) \to E$ as $t \to \infty$. If this Y and (6) are substituted into (9), then (9) defines, for large t, an $X = X(t)$ satisfying (2), which means that (3) has (*)-character. This proves the sufficiency of (i).

In order to prove the sufficiency of (ii), put $Z = X^{-1}$. Then it is readily seen from (7) that (8) can be transformed into

$$Z' = (ZA^\circ)' + ZAA^\circ. \tag{8 bis}$$

This latter equation can be written in the form $W' = ZAA^\circ$, where

$$W = Z(E - A^\circ). \tag{12}$$

But (12) and (6) show that, if t is large enough, $W' = ZAA^\circ$ is equivalent to

$$W' = W(E - A^\circ)^{-1}AA^\circ. \tag{13}$$

Finally, it is seen from (7) that (13) can be written in the form (11) by placing $Y = W^{-1}$ and

$$C = -(E - A^\circ)^{-1}AA^\circ. \tag{14}$$

Suppose that $A(t)$ satisfies condition (ii). Then, by repeating the argument applied above to the case (10) of (11), it follows that the case (14) of (11) has a solution $Y = Y(t)$ satisfying $Y(t) \to E$ as $t \to \infty$. Starting with this $Y(t)$ and applying the three definitions $W = Y^{-1}$, (12), $X = Z^{-1}$ successively (the second of them in conjunction with (6)), one

obtains an $X(t)$ satisfying (2). This proves that (ii) is sufficient for the (*)-character of $A(t)$.

It is clear from the preceding proofs that the arguments leading to the sufficiency of (i) or (ii) can be repeated so as to assume either of the conditions

(i bis) $B^\circ(t)B(t) \in L$, (ii bis) $B(t)B^\circ(t) \in L$

(along with the existence of $B^\circ(t)$, i. e., with $B(t)$) for either of the matrices $B = A^\circ A$, $B = AA^\circ$. In addition, these processes can be combined and repeated indefinitely. Thus there results an infinite scale of criteria each of which is sufficient for the (*)-character of $A(t)$.

All of these criteria assume (4), since otherwise $A^\circ(t)$ does not exist. It is therefore natural to ask whether the existence of $A^\circ(t)$ is a necessary condition for the (*)-character of $A(t)$. The answer to this question is in the negative: *$A(t)$ can be of type (*) when $A(t) \notin R$.* This will be proved by an example of an n-rowed $A(t)$, with $n = 2$. (Such an example cannot exist in the scalar case, $n = 1$, of (3), since

$$\det X(t) = \text{const.} \exp \int_{t_0}^{t} \text{trace } A(s)\,ds, \text{ where const.} \neq 0, \tag{15}$$

holds for every fundamental matrix $X(t)$ for every $n \geqq 1$.)

In order to prove the last italicized statement, let $A(t)$ be unspecified, suppose that (3) has a solution $X(t)$ satisfying (2) and put $X(t) = E + V(t)$. Then

$$V(t) = o(1), \tag{16}$$

as $t \to \infty$, and (3) shows that $A = X'X^{-1} = V'(E + V)^{-1} = V' \sum_{n=0}^{\infty} (-V)^n$ if t is large enough; cf. (16). Thus

$$A(t) = V'(t) - V'(t)V(t) + U(t), \text{ where } U = \sum_{n=2}^{\infty} (-1)^n V'V^n. \tag{17}$$

Since the formulation (16) of the (*)-character of (3) implies that $V'(t) \in R$, it follows from (17) that $A(t) \notin R$ if

$$V'(t)V(t) \notin R \tag{18}$$

but $U(t) \in R$. On the other hand, it is clear from the definition of U in (17) that $U(t) \in L$ ($\supset R$) if, as $t \to \infty$,

$$V'(t) = O(1) \text{ and } V(t) = O(t^{-1}). \tag{19}$$

Since the second of the conditions (19) implies (16), it follows that, in order to prove the existence of an $A(t) \notin R$ possessing (*)-character, it is sufficient to assure the existence of *some* (continuously differentiable) square matrix $V = V(t)$ satisfying the conditions of (16), (18), (19). In fact, any such $V(t)$ defines, by (17), an $A(t)$ possessing the desired properties.

In view of the remark made before (15), such a $V(t)$ cannot exist in the scalar case ($n = 1$). On the other hand, the binary matrix

$$(20) \qquad V(t) = \frac{1}{t}\begin{pmatrix} e(t) & 0 \\ 1/e(t) & -e(t) \end{pmatrix}, \text{ where } e(t) = \exp(-it^2),$$

is such a $V(t)$ (for $1 \leqq t < \infty$). In fact, it is readily verified from (20) that if $V'V = \| f_{\alpha\beta} \|$ then $f_{21}(t) = 4i/t$, which implies that (18) holds, while (19) is clear from (20).

The interest of this example lies in the circumstance that not even *the first step in the process of the successive approximations to* (2)-(3) can be made if (4) is violated. In fact, this process consists of the assignments

$$(21) \qquad X_{k+1}'(t) = A(t)X_k(t), X_{k+1}(\infty) = E, \text{ where } X_0(t) \equiv E.$$

But (5) shows that (21) leads for $k = 0$ to $X_1(t) = E - A^\circ(t)$ or to no $X_1(t)$ according as (4) is or is not satisfied.

Actually, not only the implicit assumption (4) of the criteria (i), (ii) but also the latter themselves can be expressed in terms of the case $k = 0$ of (21). In order to see this, suppose (4) and put

$$(22) \qquad Y(t) = E - A^\circ(t), \qquad Z(t) = (E - A^\circ(t))^{-1};$$

so that $Y = X_1$, $Z = X_1^{-1}$ (for large t; cf. (6)). These matrices are solutions of the respecitve differential equations

$$(23) \qquad Y' = B(t)Y, \qquad Z' = C(t)Z,$$

where $B = Y'Y^{-1} = (E - A^\circ)'(E - A^\circ)^{-1}$ or, by (5),

$$(24) \qquad B = A(E - A^\circ)^{-1}, \qquad C = -(E - A^\circ)^{-1}A$$

(the second of the relations (24) is a consequence of the first in view of $Z = Y^{-1}$ and of the rule mentioned before (7)). On the other hand, (6) and (22) show that (2) is satisfied by $X = Y$ and $X = Z$. In view of (23)-(24), this means that $B(t)$ and $C(t)$ are of type (*) under the assumption (4) alone.

Under the respective additional assumptions (ii), (i), it is easily concluded from (6) and (24) that

$$(25) \qquad (A(t) - B(t)) \in L, \qquad (A(t) + C(t)) \in L.$$

Accordingly, the rôle of the assumptions of the (*)-criteria (ii), (i) for $A(t)$ is that of assuring the L-character of the "small perturbations" of the coefficient matrix $A(t)$ or of the corresponding coefficient matrix $-A(t)$ which results if the solution matrix X is replaced by $Y = X_1$ or $Z = X_1^{-1}$; cf. (23). Correspondingly, since $B(t)$ and $C(t)$ in (24) are of type (*), the sufficiency of (ii) or (i) for the (*)-character of $A(t)$ could also be proved by applying to (23) the method of the variation of constants and using the fact that, according to the (*)-criterion (I), the (*)-character of $A(t) - B(t)$, $A(t) + C(t)$ is assured by (25).

(It does not matter that, by this approach to (ii), (i), the coefficient matrix $A(t)$ of (3) becomes replaced by $-A(t)$ in the second case of (25). In fact, it is clear from (4)-(5) and from the definitions of the function classes R, L that either of the conditions (i), (ii) is satisfied by $A(t)$ if and only if it is satisfied by $-A(t)$. On the other hand, since the knowledge of a solution $X(t)$ of (3) contributes no information whatever to a solution $W(t)$ of $W' = -A(t)W$, there is nothing to indicate that $-A(t)$ *must* be of type (*) when $A(t)$ is of type (*) for *some* reason.)

It will finally be shown that the existence of $X_1(t)$ is not only unnecessary but insufficient as well: $A(t)$ *need not have* (*)*-character if* $A(t) \in R$. In order to prove this converse of the negation proved above with the aid of the matrix $V(t)$, an $A(t)$ is needed which, as before, is binary ($n = 2$), at least. In fact, it is clear from (15) that condition (4) is not only necessary but sufficient as well for the (*)-character of $A(t)$ in the scalar case ($n = 1$).

Let $A(t)$ be the binary matrix in which the first and second rows are $a(t), 0$ and $b(t), -a(t)$, respectively, where

$$(26) \qquad a(t) \in R, \qquad b(t) \in R.$$

Then (4) is satisfied and (1) can be written in the form of the system $u' = a(t)u$, $v' = b(t)u - a(t)v$, where u, v denote the components of the vector x. Since, as readily verified, a solution of this system is given by

$$(27) \qquad u(t) = \exp -\int_t^\infty a(s)\,ds, \qquad v(t) = \{\int^t b(s)u^2(s)\,ds\}/u(t),$$

the last italicized assertion will be proved if it is shown that the functions

(27) need not tend to limits, as $t \to \infty$, if only (26) is assumed. But it is clear from (27) that $u(t) \to 1$ and that $v(t)$ will therefore fail to tend to a limit if $b(t)u^2(t) \notin R$, that is, if

$$b(t) \exp \int_t^\infty -2a(s)ds \notin R. \tag{28}$$

Consequently, it is sufficient to ascertain that (26) is compatible with (28).

To this end, let $b(t)$ be any (continuous) function satisfying $b(t) \in R$ and $b(t) \notin L$ (for instance, let $b(t) = (\sin t)/t$). Then it is possible to choose a (continuously differentiable) function $f(t)$ which tends, as $t \to \infty$, to 1 and is such that $b(t)f(t) \notin R$. Clearly, (26) and (28) become satisfied if $2a(t)$ is chosen to be the logarithmic derivative of such an $f(t) = f_b(t)$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] E. Hilb, Article IIB5 (1914) in the *Encyclopädie der mathematischen Wissenschaften*, vol. II_2.

[2] A. Wintner, "On linear asymptotic equilibria," *American Journal of Mathematics*, vol. 71 (1949), pp. 853-858.

[3] ———, "On free vibrations with amplitudinal limits," *Quarterly of Applied Mathematics*, vol. 8 (1950), pp. 102-104.

COMMUTATORS OF OPERATORS, II.*

By Paul R. Halmos.

1. Introduction. Suppose that A and B are linear transformations on a complex Hilbert space H and let C be their commutator, $C = AB - BA$. Problems concerning C can be studied on three levels of generality. On the lowest level, it is assumed that H is finite-dimensional; on the intermediate level, infinite-dimensional spaces are allowed, but only bounded linear transformations (here called *operators*) are considered; and, on the highest level, both the space and the transformations remain completely unrestricted. Since, roughly speaking, on the highest level anything can happen, whereas on the lowest level nothing can happen, the mathematically interesting and difficult problems are all likely to be on the intermediate level. Thus, for example, the identity operator is a commutator of unbounded transformations, but not of operators, and the identity operator is the sum of two commutators of operators, whereas the identity matrix (of finite size) is not the sum of any finite number of commutators of matrices.

The purpose of this note is to obtain some positive and some negative results concerning commutators on the middle level, and to call attention to some unsolved problems. The results are of three types. Section 2 studies the role which 0 plays in the spectrum of a commutator. Section 3 concentrates on the problem of approximation by commutators, and, in particular, on how near a commutator can be to the identity. In Section 4 a factorization theorem, valid for every operator, is proved and applied to the problem of studying the spectrum of a commutator.

The three levels of operatorial generality have their analogs in pure algebra; the lowest level corresponds to finite-dimensional algebras, the middle level to normed algebras, and the highest level to completely unrestricted algebras. At times the relation becomes more than an analogy; in particular, it will be seen that several of the proofs below can be transferred from (bounded) operators to (normed) algebras without any change. Nevertheless, in order to avoid unnecessary complications, the more special terminology and notation of operator theory will be preferred throughout.

It is occasionally convenient to use the differential notation for commu-

* Received October 1, 1953.

tators. Thus, to study $AB - BA$, the operator B will be fixed once and for all, and the commutator, regarded as a function of A, will be denoted by A'. The mapping $A \to A'$ is a derivation; i. e., it is a linear transformation from operators to operators that has many of the properties of ordinary differentiation. Thus

$$(A_1A_2)' = A_1A_2B - BA_1A_2$$
$$= A_1(A_2B - BA_2) + (A_1B - BA_1)A_2 = A_1A'_2 + A'_1A_2,$$

and therefore the non-commutative versions of all the usual Leibniz formulas are valid. The product formula applied to A^k $(k = 1, 2, \cdots)$ yields $(A^k)' = A^{k-1}A' + (A^{k-1})'A$; by an obvious induction argument it follows that

$$(A^k)' = \sum_{i=0}^{k-1} A^{k-1-i}A'A^i.$$

2. Divisors of zero. In the algebraic study of commutators an important role is played by Jacobson's lemma ([4], Lemma 2). That lemma asserts that, in an algebraic algebra over a field of characteristic zero, if A' commutes with A, then A' is nilpotent. Kaplansky conjectured that an analog of this result is valid for normed algebras, i. e., that if A' commutes with A, then A' is quasi-nilpotent. (Recall that, by definition, A' is quasi-nilpotent if $\| (A')^n \|^{1/n} \to 0$ as $n \to \infty$, or, equivalently, if the spectrum of A' consists of 0 alone. Still another formulation says that $r(A') = 0$, where r, here and below, denotes the spectral radius. For a discussion of such concepts and their role in normed algebras see [3], Chapter V.) Kaplansky's conjecture is not settled yet; the following results, however, are corroborative evidence.

THEOREM 1. *If A' commutes with A, then A' is a (two-sided) generalized divisor of zero.*

Proof. The equation for the "derivative" of A^n implies that under the present commutativity hypothesis $(A^n)' = nA^{n-1}A'$. It follows that if A is nilpotent (in the usual algebraic sense), then A' is a divisor of zero (also in the algebraic sense). If A is not nilpotent, then the relation

$$n \| A^{n-1}A' \| = \| A^nB - BA^n \| \leqq 2 \| A \| \cdot \| B \| \cdot \| A^{n-1} \|$$

implies that $\| A^{n-1}A' \| / \| A^{n-1} \| \to 0$ as $n \to \infty$, i. e., that A' annihilates the

sequence $\{A^{n-1}/\| A^{n-1} \|\}$ of operators of norm 1. Since this is the definition of a generalized divisor of zero, the proof is complete.[1]

The above proof is valid in any normed algebra. The method was first used by Wielandt in [6] to prove that A' cannot be equal to 1; for an earlier and different proof, see Wintner [7]. That result, of course, is an immediate consequence of Theorem 1. Equally immediate is the following corollary, which is being put on record here for convenience of reference.

COROLLARY. *If A' commutes with A, then A' is not invertible.*

The conclusion of Theorem 1 is expressed in normed-algebraic terms; it is sometimes convenient to be able to use operatorial language. This desideratum is achieved by the following slight weakening of Theorem 1.

THEOREM 2. *If A' commutes with A, then 0 belongs to the approximate point spectrum of A'.*

(For related results, see [5], especially Theorem II, p. 359.)

Proof. It is sufficient to prove that an operator C is a generalized left divisor of zero if and only if 0 belongs to its approximate point spectrum. Suppose therefore that $CD_n \to 0$, with $\| D_n \| = 1$, and let x_n be a vector such that $\| D_n x_n \| \geqq \frac{1}{2} \| x_n \|$. If $y_n = D_n x_n / \| D_n x_n \|$, then $\| y_n \| = 1$, and

$$\| Cy_n \| = \| CD_n x_n \| / \| D_n x_n \| \leqq \| CD_n \| \cdot \| x_n \| / (\tfrac{1}{2}) \| x_n \| = 2 \| CD_n \| \to 0.$$

If, conversely, $Cx_n \to 0$ as $n \to \infty$, with $\| x_n \| = 1$, write $D_n x = (x, x_n) x_n$. It follows that D_n is the projection on the one-dimensional space spanned by x_n, so that $\| D_n \| = 1$. The relation

$$\| CD_n x \| = \| C(x, x_n) x_n \| \leqq \| Cx_n \| \cdot \| x \|,$$

valid for all n and x, implies that $\| CD_n \| \leqq \| Cx_n \| \to 0$ as $n \to \infty$.

3. Density. The results of the preceding section are essentially negative; since they say that commutators must satisfy certain conditions, their chief use is in proving that certain operators are not commutators. In this section and the next the emphasis will be positive; it will be shown that on an infinite-dimensional Hilbert space there are many commutators.

LEMMA 1. *If A is an operator on a Hilbert space H, and if M is a*

[1] I should like to express my appreciation to Günter Lumer and Juan J. Schäffer for many interesting conversations on the subject of commutators. Theorem 1, in particular, was obtained in collaboration with Schäffer.

subspace of H such that $\dim M \leqq \dim M^{\perp}$, *then there exists a commutator C on H such that $Cx = Ax$ whenever $x \in M$.*

Proof. Let Q be a partial isometry with initial space M and final space $N \subset M^{\perp}$. In other words, Q sends M isometrically onto N and annihilates $M^{\perp}$, whereas Q^* sends N isometrically onto M and annihilates $N^{\perp}$ (and therefore annihilates M). If $P = AQ^*$, then the desired commutator is $C = PQ - QP$; indeed, if $x \in M$, then

$$Cx = PQx - QPx = AQ^*Qx - QAQ^*x = Ax.$$

THEOREM 3. *If H is infinite-dimensional, then the set of commutators is dense, with respect to the strong topology, in the set of all operators on H.*

Proof. For any operator A and vectors $x_1, \cdots, x_n$, let M be the space spanned by the x's, and apply Lemma 1 to obtain a commutator C such that $Cx_j = Ax_j$, $j = 1, \cdots, n$. The existence of such a C implies that every strong neighborhood of A contains a commutator.

The preceding result makes it natural to ask whether or not the set of commutators is dense with respect to the uniform (norm) topology as well. The answer to that question is not known. In view of this ignorance, and in view of the quantum-mechanical history of the study of commutators, it becomes pertinent to ask: how well can the identity operator be approximated by commutators? In finite-dimensional spaces the answer is easy.

THEOREM 4. *If H is finite-dimensional, then $r(1 - A') \geqq 1$ for all A.*

Proof. For any operator A on H, let $t(A)$ denote the normalized trace of A, i. e., the average of the proper values of A. Since $t(A') = 0$ for all A, it follows that not all the proper values of A' can lie in the circle $\{\lambda : |1 - \lambda| < 1\}$.

COROLLARY. *If H is finite-dimensional, then $\| 1 - A' \| \geqq 1$ for all A.*

Proof. This follows immediately from Theorem 4, together with the fact that the normalized trace is dominated by the norm.

Note that, that since 0 is a commutator, the result of the corollary is best possible.

The main purpose of the next result is purely negative; it shows that the hypothesis of finite-dimensionality is indispensable in the corollary to Theorem 4.

THEOREM 5. *If H is infinite-dimensional, then there exists a commutator C on H such that $\| 1 - C \|^2 \leqq .97$.*

Proof. Let U and V be isometries on H with respective final spaces M and N, where $M = N^\perp$. (The existence of U and V is ensured by the fact that H is infinite-dimensional.) In other words, $U^*U - UU^* = 1 - UU^*$ is the projection on $M^\perp$, and $V^*V - VV^* = 1 - VV^*$ is the projection on $N^\perp = M$; it follows that $U^*U - UU^* + V^*V - VV^* = 1$. If P and Q are defined by

$$2^{\frac{1}{2}}P = (U + U^*) + i(V + V^*), \qquad 2^{\frac{1}{2}}Q = (U - U^*) - i(V - V^*),$$

then a straightforward computation shows that $C = PQ - QP = 1 + iR$, where R is the Hermitian operator $(U^*V^* - V^*U^*) - (UV - VU)$. It follows that $\| R \| \leqq 4$. (In my earlier paper [1] I showed that every operator, and in particular the identity, can be the Hermitian part of a commutator. Up to this point the present proof succeeded only in re-establishing that result, in a somewhat simpler and more explicit form than before, and in furnishing an estimate for the size of $\| R \|$.)

Every element of the spectrum of C is of the form $1 + i\alpha$, where α is a real number such that $| \alpha | \leqq \| R \| \leqq 4$. If δ is a positive number, and if λ belongs to the spectrum of δC, then $\lambda = \delta + i\delta\alpha$, so that

$$| 1 - \lambda |^2 = (1 - \delta)^2 + \delta^2\alpha^2.$$

It follows that if $\delta = .1$, then $| 1 - \lambda |^2 \leqq .97$. Three comments complete the proof. First: since the Hermitian part of C is 1, and therefore commutative with its skew-Hermitian part, the operator C is normal (and so also is δC). Second: since, for a normal operator, the spectral radius and the norm coincide, $\| 1 - \delta C \|^2 \leqq .97$. Third: since C is a commutator, so is δC.

In general the infimum of $\| 1 - C \|$, as C varies over all commutators, is not known. There are, however, two special situations in which the inequality of the corollary to Theorem 4 remains valid.

THEOREM 6. *If A' commutes with A, then $\| 1 - A' \| \geqq 1$.*

Proof. If $\| 1 - A' \| < 1$, then A' is invertible (see [3], Chapter V); the desired result follows from the corollary to Theorem 1.

THEOREM 7. *If the operator A is such that*

$$\limsup n^{-1} \sum_{i=0}^{n-1} \| A^{n-1-i} \| \cdot \| A^i \| / \| A^{n-1} \| \leqq 1,$$

then $\| 1 - A' \| \geqq 1$.

Remark. Since

$$n = \sum_{i=0}^{n-1} \| A^{n-1-i} \cdot A^i \| / \| A^{n-1} \| \leqq \sum_{i=0}^{n-1} \| A^{n-1-i} \| \cdot \| A^i \| / \| A^{n-1} \|,$$

the hypothesis trivially implies that the limit exists and is equal to 1.

Proof. The Wielandt trick can be applied again. Since

$$(A^n)' = \sum_{i=0}^{n-1} A^{n-1-i} A' A^i = nA^{n-1} \sum_{i=0}^{n-1} A^{n-1-i}(1 - A')A^i,$$

it follows that

$$n \| A^{n-1} \| \leqq \| A^n B - BA^n \| + \| \sum_{i=0}^{n-1} A^{n-1-i}(1 - A')A^i \|$$

$$\leqq 2 \| A \| \cdot \| B \| \cdot \| A^{n-1} \| + \| 1 - A' \| \sum_{i=0}^{n-1} \| A^{n-1-i} \| \cdot \| A^i \|.$$

Division by $\| A^{n-1} \|$ yields

$$n \leqq 2 \| A \| \cdot \| B \| + \| 1 - A' \| \sum_{i=0}^{n-1} \| A^{n-1-i} \| \cdot \| A^i \| / \| A^{n-1} \|.$$

It follows that the hypothesis of the theorem is not consistent with the relation $\| 1 - A' \| < 1$, and the proof is complete.

The hypothesis of Theorem 7 looks weird; its principal application is to subnormal (and a fortiori to normal) operators. An operator is subnormal if it has a normal extension (to a larger Hilbert space). The point is that if A is subnormal, then $\| A^k \| = \| A \|^k$ for every positive integer k. (For normal operators this is well known; for subnormal operators, it follows from the results of [2].) In view of these comments, and of the fact that an isometry is easily seen to be subnormal, the following assertion is an immediate consequence of Theorem 7.

COROLLARY. *If A is subnormal (and, in particular, if A is an isometry, or if A is normal), then* $\| 1 - A' \| \geqq 1$.

The considerations of this section have been guided mainly by the question: can a commutator be near to the identity in the sense of norm? A related question is: can a commutator C be near to the identity in the purely algebraic sense that $1 - C$ is nilpotent? In the finite-dimensional case the answer is obviously no; the trace of a nilpotent operator is 0, whereas the (normalized) trace of $1 - C$ is 1. In the general case this problem takes its place with the many other problems concerning commutators of operators that still remain unsolved.

4. Factorization. The main result (Theorem 8) of this final section is somewhat paradoxical; it asserts that (if H is infinite-dimensional) every operator behaves in some respects like a commutator.

If K is the direct sum of two copies of H, then every operator on K can be considered as a two-rowed matrix whose entries are operators on H. The following auxiliary result considers such matrices and proves that certain special ones among them are commutators. Since, in the infinite-dimensional case, H and K are isomorphic, the result could also be formulated in terms of H alone, but only at a considerable cost in clarity.

LEMMA 2. *If H is infinite-dimensional, then to every operator A on H there correspond two matrices*

$$X = \begin{pmatrix} 0 & P \\ 0 & R \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 \\ Q & S \end{pmatrix} \text{ with } \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} = XY - YX,$$

where P, Q, R, and S are operators on H.

Proof. If K is the direct sum of a countably infinite number of copies of H, then every operator on K can be written as an infinite matrix whose entries are operators on H. Just as in the case of numerical (infinite) matrices, not every such matrix corresponds to an operator. If, however, X is the matrix whose entries in the diagonal just above the main diagonal are all equal to A, and if Y is the matrix whose entries in the diagonal just below the main diagonal are all equal to 1, all other entries in both cases being 0, then X and Y do correspond to operators on K. It is easy to verify that the commutator $XY - YX$ has A in the $(1, 1)$ position and 0 everywhere else. Since the first "axis" of K is isomorphic to the direct sum of all the others, i. e., since K may also be considered as the direct sum of two copies of H, the result just obtained can also be expressed in terms of two-rowed matrices. The effect of expressing it in such terms is exactly the desired conclusion.

THEOREM 8. *Every operator on an infinite-dimensional space can be written as the product of two operators whose product in the reverse order is a commutator.*

Proof. An application of Lemma 2 to the prescribed operator A shows the existence of operators P, Q, R, and S such that

$$\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} PQ & PS \\ RQ & (RS - SR) - QP \end{pmatrix}.$$

It follows that $A = PQ$ and $QP = RS - SR$.

COROLLARY. *Every operator on an infinite-dimensional space is the sum of two commutators.*

Proof. If $A=PQ$ with $QP=RS-SR$, then $A=PQ-QP+RS-SR$.

This corollary is an improvement of one of the results of [1]; the methods of [1] yielded the same statement with *four* in place of *two*.

What can the spectrum of a commutator look like? Can it be an arbitrary compact subset of the complex plane? Here again the answer is not known. Easy finite-dimensional examples show that 0 need not belong to the spectrum of a commutator. If, however, 0 does belong to the spectrum, then the rest of the structure of the spectrum can be prescribed arbitrarily. The precise formulation of this somewhat vague statement is the following application of Lemma 2.

THEOREM 9. *If Λ is a compact subset of the complex plane, with $0 \in \Lambda$, then there exists a commutator C whose spectrum is Λ.*

Proof. Let A be an operator whose spectrum is Λ; the existence of an operator, and even of a normal operator, with prescribed spectrum is a standard exercise in elementary Hilbert space theory. If $C=\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$, then, since $0 \in \Lambda$, the spectrum of C is also Λ; the desired result follows from Lemma 2.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] P. R. Halmos, "Commutators of operators," *American Journal of Mathematics*, vol. 74 (1952), pp. 237-240.

[2] ———, "Spectra and spectral manifolds," *Annales de la Société Polonaise de Mathématique*, vol. 25 (1952), pp. 43-49.

[3] E. Hille, *Functional analysis and semi-groups*, New York, 1948.

[4] N. Jacobson, "Rational methods in the theory of Lie algebras," *Annals of Mathematics*, vol. 36 (1935), pp. 875-881.

[5] C. R. Putnam, "On normal operators in Hilbert space," *American Journal of Mathematics*, vol. 73 (1951), pp. 357-362.

[6] H. Wielandt, "Ueber der Unbeschränktheit der Operatoren der Quantenmechanik," *Mathematische Annalen*, vol. 121 (1949), p. 21.

[7] A. Wintner, "The unboundedness of quantum-mechanical matrices," *Physical Review*, vol. 71 (1947), pp. 738-739.

LINEAR DIFFERENTIAL EQUATIONS WITH COMPLETELY MONOTONE SOLUTIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

---

1. Let the coefficient $f(x)$ in a linear differential equation of the form

$$y'' - f(x)y = 0 \tag{1}$$

be completely monotone on the half-line $(0,\infty)$, that is, let all derivatives $f^{(k)}(x)$ of $f(x)$ exist and satisfy

$$(-1)^k f^{(k)} \geqq 0 \qquad (k = 0, 1, \cdots) \tag{2}$$

on $(0,\infty)$. It was shown in [5] that (1) must then possess a solution $y(x) \not\equiv 0$ which is completely monotone on $(0,\infty)$. In view of the Hausdorff-Bernstein theorem, this means that, if the conditions (2) (which in general can readily be checked for the coefficient function of (1)) are satisfied, then (1) must possess a solution which is representable on $(0,\infty)$ in the form

$$y(x) = \int_0^\infty e^{-xt} d\mu(t), \qquad \text{where } 0 \leqq d\mu(t) \not\equiv 0. \tag{3}$$

Thus the existence of a Laplace solution (in fact, of one having the particular structure specified by (3)) follows without any use of the classical technique of contour integrations, which can be applied (and is then laborious indeed) only if f is given explicitly. Correspondingly, the theorem, being a mere existence statement, can supply only the existence, but no indication leading to an explicit determination, of the function μ occurring in (3).

(The theorem does have some "explicit" use, but one in the reverse direction, namely, as follows: If f satisfies (2) on $(0,\infty)$ and if it is known that (1) possesses a solution $y(x)$ which remains bounded as $x \to \infty$ and is positive at some $x = x_0 > 0$, then $(-1)^k$ times the k-th derivative of $y(x)$ must be non-negative for every $k \geqq 0$ and for every $x > 0$. For a simple proof, cf. [6], where an application is given to Whittaker's function W_{mn}.)

---

* Received August 18, 1953.

2. The existence theorem, mentioned at the beginning of Section **1**, was extended in [1] from (1) to

$$h(x)y'' + g(x)y' - f(x)y = 0 \tag{4}$$

under the following assumptions: On $(0,\infty)$,

$$h > 0 \tag{5*}$$

and

$$h', g, f \text{ are completely monotone.} \tag{5}$$

In other words, (5*) and (5), where $h' = dh/dx$, assure for the homogeneous, linear differential equation (4) the existence of a solution of the form (3). Needless to say, if (5_h), (5_g), (5_f) denote the respective assumptions in (5), then (5_f) is identical with (2) and (5*), (5_h), (5_g) are satisfied in the particular case (1) of (4).

If $\alpha > 0$ in Kummer's form of the confluent hypergeometric equation,

$$xy'' + (\gamma - x)y' - \alpha y = 0, \tag{6}$$

or in the closely related differential equation of the parabolic cylinder,

$$y'' - xy' - \alpha y = 0, \tag{7}$$

then the well-known integrals

$$y(x) = \int_0^\infty e^{-xt} t^{\alpha-1}(1+t)^{\gamma-\alpha-1}dt \qquad (0 < x < \infty) \tag{6 bis}$$

or

$$y(x) = \int_0^\infty e^{-xt} t^{\alpha-1} \exp(-\tfrac{1}{2}t^2)dt \qquad (-\infty < x < \infty), \tag{7 bis}$$

which are of the form (3), are solutions (the existence of solutions (3) in the limiting cases $\alpha = 0$ is trivial, since $y(x) \equiv 1$, which belongs to $\mu(t) = \operatorname{sgn} t$ in (3), is then a solution). But the existence of these classical solutions is not accounted for by the criterion (5*)-(5). In fact, the circumstance that the coefficients, $g(x) = \gamma - x$ and $g(x) = -x$, of the cases (6), (7) of (4) become negative (for large x) prevent the satisfaction of (5*)-(5), if $\alpha \geqq 0$.

3. While these cases of (4) violate (5_g), they satisfy (5*) and are such that

$$h', -g', f \text{ are completely monotone.} \tag{8}$$

Hence, what is at stake is whether (5_g) can be generalized to (8_g) in the sufficient criterion (5*)-(5) for the existence of a solution (3) of (4) (that (8_g) is a generalization of (5_g) follows from the fact that the negative of the first derivative of a completely monotone function is completely monotone, while the converse is not true).

It will be proved that the answer to this question is affirmative. In addition, condition (8_h) can be replaced by the assumption that $-h''$ is completely monotone. Actually, since (5*) is retained and since

(9) $\qquad h > 0$ and $h'' \leqq 0$ imply that $h' \geqq 0$, if $0 < x < \infty$,

this is no further generalization of (5). In other words, (8) is equivalent to the assumptions that

(10) $\qquad -h'', -g', f$ are completely monotone, $\qquad (0 < x < \infty)$,

if (5*) is satisfied. Accordingly, the theorem to be proved can be formulated as follows:

(*) *If the coefficients of* (4) *satisfy the assumptions* (10) *and* (5*) *on* $(0,\infty)$, *then* (4) *must possess a solution* $y(x)$ *which is representable on* $(0,\infty)$ *in the form* (3).

Needless to say, if $y(x)$ is such a solution, then so is any positive constant multiple of $y(x)$. Except for this trivial indeterminacy, the solution (3) is necessarily unique if the coefficients of (4) violate certain conditions. The latter follow in an explicit form from a more general result of [2], p. 382 (a result which assumes much less than (10) and (5*)).

4. The theorem announced in Section **3** is contained in the following theorem (**). If $n = 2$, then (**) reduces to (*). For an arbitrary n, that particular case of (**) which corresponds to the particular case (5) of (8) is known; cf. (i bis) in [3], pp. 731-732.

(**) *Let the half-line* $(0,\infty)$ *be the range of* x *in the differential equation*

(11) $$f_0(x)D^n y + \sum_{k=1}^{n} (-1)^{k-1} f_k(x) D^{n-k} y = 0,$$

where $D = d/dx$ *and* $n \geqq 2$, *and suppose that*

(12) $\qquad -D^2 f_0, -Df_1$ and $f_2, \cdots, f_n$ are completely monotone

and

(13) $$f_0 > 0.$$

Then (11) *possesses a solution* $y(x)$ *which is representable on* $(0,\infty)$ *in the form* (3).

5. The proof of (**) will depend on the following

LEMMA. *In* (11), *let* $n > 2$, *and let* (13) *and*

$$f_j(x) \geqq 0 \qquad (j = 2, 3, \cdots, n) \tag{14}$$

hold on $0 < x < \infty$ (*while* $f_1 \gtrless 0$). *Then, if a solution* $y = y(x) \not\equiv 0$ *of* (11) *satisfies*

$$(-1)^k D^k y \geqq 0 \qquad (0 < x < \infty) \tag{15_k}$$

for $k = 0, 1, \cdots, n-2$, *it must satisfy* (15_k) *for* $k = n-1$ *also.*

In order to verify this lemma, note that (11) and (15_k), where $k = 0, 1, \cdots, n-2$, imply that $(-1)^n(f_0 D^n y + f_1 D^{n-1} y) \geqq 0$; hence

$$D\{(-1)^{n-1}(D^{n-1}y)\exp\int_X^x (f_1/f_0)\,ds\} \leqq 0$$

for $0 < X \leqq x < \infty$. If (15_{n-1}) fails to hold for some x-value, say for $x = X > 0$, then $(-1)^{n-1} D^{n-1} y < 0$ for $X \leqq x < \infty$. Consequently, $(-1)^{n-1} D^{n-1} y$ is decreasing for $X \leqq x < \infty$; so that $(-1)^{n-2} D^{n-2} y \geqq 0$ is increasing for $X \leqq x < \infty$. In particular

$$(-1)^{n-2} D^{n-2} y \geqq \text{const.} > 0 \text{ for large } x. \tag{16}$$

On the other hand, $n \geqq 3$ and the inequalities (15_{n-3}), (15_{n-2}) imply that $D^{n-3}y$ tends to a (finite) limit as $x \to \infty$. But a quadrature shows that this is incompatible with (16), that is, with the assumption that (15_{n-1}) can fail to hold for some value of $x = X > 0$.

6. The proof of (**) proceeds as follows:

The assumptions of (**) imply (13) and (14). Hence, assertion (i) of [3] is applicable and assures the existence of a solution $y = y(t) \not\equiv 0$ of (11) satisfying (15_k), where $k = 0, 1, \cdots, n-1$. It will be shown, by induction, that this solution satisfies (15_k) for $k = 0, 1, \cdots$.

Assume that (15_k) holds for $k = 0, 1, \cdots, n+j-2$ for some $j \geqq 1$. It will be shown that it also holds for $k = n+j-1$. To this end, let (11) be differentiated j times. The resulting equation,

$$\sum_{m=0}^{j} C_{jm}\{(D^m f_0)(D^{n+j-m}y) + \sum_{k=1}^{n}(-1)^{k-1}(D^m f_k)(D^{n+j-k-m}y)\} = 0,$$

where $C_{jm} = j!/m!(j-m)!$, can be written in the form

$$f_0(x)D^{n+j}y + \sum_{k=1}^{n+j}(-1)^{k-1}F_k(x)D^{n+j-k}y = 0, \tag{17}$$

where $F_k = F_{jk}(x)$ is defined by

$$F_k = (-1)^{k-1}C_{jk}D^k f_0 + \sum_{m=0}^{k-1}(-1)^m C_{jm}D^m f_{k-m} \tag{18}$$

when $j = 1, 2, \cdots$ and $k = 1, 2, \cdots, j$, and by

$$F_k = \sum_{m=0}^{j}(-1)^m C_{jm}D^m f_{k-m} \tag{19}$$

when $j = 1, 2, \cdots$ and $k = j+1, j+2, \cdots, j+n$.

The assumptions (12) imply that

$$(-1)^{k+1}D^{k+2}f_0 \geqq 0, \quad (-1)^k D^{k+1}f_1 \geqq 0 \text{ and } (-1)^k D^k f_j \geqq 0 \tag{20}$$

for $k = 0, 1, \cdots$ and $j = 2, \cdots, n$. Hence, (13) and 20) show that

$$F_k(x) \geqq 0 \text{ for } k = 2, 3, \cdots, n+j, \tag{21}$$

since neither Df_0 nor f_1 occurs in the formulae for F_k when $k \geqq 2$. By the induction hypothesis, the solution $y = y(x)$ of (17) satisfies (15_k) for $k = 0, 1, \cdots, n+j-2$. Hence, the Lemma of Section **5** shows that (13) and (21) imply that $y = y(x)$ satisfies (15_k) for $k = n+j-1$ (note that the condition $n > 2$ of the Lemma is satisfied by the present $n + j \geqq 2 + j > 2$). This completes the induction and the proof of (**).

7. Both in the assumptions and the assertion of (**), the x-range was the open half-line $(0, \infty)$. It is however clear that the proof applies on the closed half-line $[0, \infty)$ also. The resulting variant of (**) can be combined with the following remark (which has nothing to do with differential equations):

If a function $y(x)$ is completely monotone on $[0, \infty)$, then the same is true of the functions $y_l(x)$, where $l = 1, 2, \cdots$ and

$$y_l(x) = \{y(x) - \sum_{k=0}^{l-1} y^{(k)}(0)x^k/k!\}/(-x)^l.$$

In order to see this, it is sufficient to observe that, according to Taylor's

formula with integral remainder, the expression $\{\cdots\}$ occurring in the definition of $y_l(x)$ is

$$(-x)^l \int_0^1 (-1)^l y^{(l)}(xt)(1-t)^{l-1}dt/(l-1)!.$$

Appendix.

Theorem (i) of [3], p. 731, on which the proof of (**) was based above, states that the conditions (13)-(14), where $0 < x < \infty$, assure the existence of a solution $y = y(x)$ satisfying (15_k) for $k = 0, 1, \cdots, n-1$. There is no condition on $f_1 (\gtreqless 0)$. Correspondingly, if $f_1 \geqq 0$ is assumed, then (15_k) holds for $k = n$ also (cf. *loc. cit.*).

It is natural to ask whether or not the assumptions (14) can be omitted for some values of j on the range $2 \leqq j \leqq n$ if, in the corresponding assertion, some values on the range $0 \leqq k \leqq n-1$ are omitted for the index k in (15_k). A partial answer is given by (†) below.

If $y_1, y_2, \cdots, y_m$ are m functions, each having m derivatives, let $W_0 \equiv 1$, $W_1 = y_1(x)$ and, if $k = 2, \cdots, m$, let $W_k = W_k(x) = W_k(x; y_1, y_2, \cdots, y_k)$ denote the (k-rowed) Wronskian determinant of the functions $y_1, \cdots, y_k$.

(†) *Let $0 < x < \infty$ and $0 \leqq m \leqq n$. In the differential equation* (11), *let*

$$f_k(x) \geqq 0 \text{ for } k = m+1, \cdots, n \tag{22}$$

and let the differential equation

$$f_0(x)D^m y + \sum_{k=1}^{m} (-1)^{k-1} f_k(x) D^{m-k} y = 0 \tag{23}$$

possess a complete set of solutions $y_1(x), \cdots, y_m(x)$ satisfying

$$W_k = W_k(x) = W_k(x; y_1, \cdots, y_k) > 0 \text{ for } k = 0, 1, \cdots, m. \tag{24}$$

Then (11) *has a solution $y = y(x)$ satisfying (15_k) for $k = 0, 1, \cdots, n-m$.*

This theorem (†) will be deduced from Theorem (I) in [3], pp. 733-734, which deals with a linear, homogeneous system of first order differential equations.

First, if $a_0(x), \cdots, a_m(x)$ denote the positive functions $a_0 = W_0/W_1$, $a_k = W_k^2/W_{k-1}W_{k+1}$, where $k = 1, \cdots, m-1$, and $a_m = W_m/W_{m-1}$, then, according to Frobenius, the expression in the right of (23) can be "factored" as

$$f_0 a_m D a_{m-1} D a_{m-2} \cdots D a_1 D a_0 y,$$

where, for example, $Da_1Da_0y = D(a_1(D(a_0y)))$ (for a simple proof, see [4], pp. 315-316). If y is replaced by $D^{n-m}y$, then the expression in the last formula line becomes identical with the first $m+1$ terms of (11). Hence, (11) can be written as

$$(25) \qquad f_0a_mDa_{m-1}D\cdots a_1Da_0D^{n-m}y + \sum_{k=m+1}^{n}(-1)^{k-1}f_kD^{n-k}y = 0.$$

Define n functions $z_1, \cdots, z_n$ by the equations

$$(26_1) \qquad z_k = (-1)^{k-1}D^{k-1}y \text{ if } k = 1, \cdots, n-m,$$

$$(26_2) \qquad z_{n-m+k} = (-1)^{n-m+k-1}a_{k-1}Da_{k-2}\cdots Da_0D^{n-m}y \text{ if } k = 1, \cdots, m.$$

Then the n-th order differential equation (11) and/or (25) is equivalent to the following system of n first order equations:

$$(27_1) \qquad z_k' = -z_{k+1} \text{ if } k = 1, \cdots, n-m-1,$$

$$(27_2) \qquad z_{n-m+k-1}' = -z_{n-m+k}/a_{k-1} \text{ if } k = 1, \cdots, m,$$

$$(27_3) \qquad z_n' = -\sum_{k=m+1}^{n} f_kz_{n-k+1}/f_0a_m.$$

Since $a_0, \cdots, a_m$ are positive, it follows from (22) that the coefficients of $-z_1, \cdots, -z_n$ on the right side of (27_1)-(27_3) are non-negative (in fact, the coefficients are $0, 1, 1/a_0, \cdots, 1/a_{m-1}, f_{m+1}/f_0a_m, f_{m+2}/f_0a_m, \cdots, f_n/f_0a_m$). Hence Theorem (I) of [3], pp. 733-734, is applicable and implies that the system (27_1)-(27_3) has a non-trivial solution $z_1 = z_1(x), \cdots, z_n = z_n(x)$ satisfying $z_k(x) \geqq 0$, where $k = 1, \cdots, n$ and $0 < x < \infty$. The definition (26_1) of z_k for $k = 1, \cdots, n-m$ and the definition of z_{n-m+1} in (26_2) now show that (11) has a solution $y = y(x) \not\equiv 0$ satisfying (15_k) for $k = 0, 1, \cdots, n-m$. This proves (†).

It may be mentioned that while Theorem (i) of [3] can readily be deduced from (†), the converse deduction is impossible. This is shown by the example

$$(28) \qquad D^3y + y = 0,$$

which has the general solution

$$(29) \qquad y = ae^{-x} + be^{\frac{1}{2}x}\cos(3^{\frac{1}{2}}x/2) + ce^{\frac{1}{2}x}\sin(3^{\frac{1}{2}}x/2).$$

In fact, the conditions imposed by (i) of [3] on (23) are satisfied by the case (28) of (11), where $m = 3$, but it is readily verified that for no choice of the integration constants a, b, c will there result $m = 3$ solutions y_1, y_2, y_3 satisfying the conditions (24).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "On the Laplace-Fourier transcendents," *American Journal of Mathematics*, vol. 71 (1949), pp. 367-372.

[2] ——— and A. Wintner, "On the classical transcendents of mathematical physics," *ibid.*, vol. 73 (1951), pp. 381-389.

[3] ——— and A. Wintner, "Linear differential and difference equations with monotone solutions," *ibid.*, vol. 75 (1953), pp. 731-743.

[4] G. Pólya, "On the mean-value theorem corresponding to a given linear homogeneous differential equation," *Transactions of the American Mathematical Society*, vol. 24 (1922), pp. 312-324.

[5] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

[6] ———, "On the Whittaker functions $W_{km}(x)$," *Journal of the London Mathematical Society*, vol. 25 (1950), pp. 352-353.

ON NON-OSCILLATORY LINEAR DIFFERENTIAL EQUATIONS WITH MONOTONE COEFFICIENTS.*

By PHILIP HARTMAN and AUREL WINTNER.

---

1. *Introduction.* Let the coefficient function in the differential equation

$$y'' + q(t)y = 0, \tag{1}$$

where $0 \leqq t < \infty$, be continuous and *monotone* (both cases

$$dq \geqq 0, \tag{2$_+$}$$

$$dq \leqq 0 \tag{2$_-$}$$

will be considered). Along every solution $y = y(t) \not\equiv 0$ of (1), define $h = h(t)$ by

$$h = y'^2 + qy^2 \tag{3}$$

and $k = k(t)$ by

$$k = y^2 + y'^2/q \quad \text{if} \quad q \neq 0. \tag{4}$$

Then, in the sense of Stieltjes integrations (in t),

$$dh = y^2 dq \tag{5}$$

and

$$dk = -y'^2 dq/q^2 \qquad (q \neq 0); \tag{6}$$

cf. e. g., [10], pp. 384-385 and [5], pp. 530-531, respectively. Hence the monotony of q implies that of both functions (3), (4) of t; in fact,

$$dh \geqq 0, dk \leqq 0, \tag{7$_+$}$$

$$dh \leqq 0, dk \geqq 0 \tag{7$_-$}$$

in the respective cases (2$_+$), (2$_-$).

On any interval on which $y(t) \neq 0$, the logarithmic derivative

$$r(t) = y'(t)/y(t) \tag{8}$$

satisfies the Riccati equation

$$r' + r^2 + q = 0 \tag{9}$$

(even if q is not monotone).

---

* Received July 9, 1953.

This paper will deal with the behavior of the functions (3), (4) and (8) when $q(t)$ is monotone and

(10) (1) is non-oscillatory.

By (10) is meant that some (hence, according to Sturm, every) solution $y(t) \not\equiv 0$ of (1) has only a finite number of zeros on the half line $0 \leqq t < \infty$. For other results on the case of (1) for which (2$_-$), (10) and $q > 0$ hold, cf. [5].

2. *Positive* $q(t)$. The first assertions will deal with the case in which q is monotone, positive and satisfies (10). Then (2$_+$) is impossible, hence (2$_-$) holds. In addition,

$$0 < q \to 0 \quad \text{as} \quad t \to \infty \tag{11}$$

and, by (10) and $q > 0$,

$$\int^{\infty} q(t)\,dt < \infty \tag{12}$$

must hold (cf. [11], p. 97), which implies, by (2$_-$), that

$$tq(t) \to 0 \text{ as } t \to \infty. \tag{13}$$

The first theorem, in contrast to most of the other theorems in this paper, will not assume that q is monotone.

(*) *If* (1) *is non-oscillatory and* $q(t) \geqq 0$, *and if* $y = y(t)$ *is any solution* ($\not\equiv 0$) *of* (1), *then, from a certain* t_0 *onward, the derivative of* (8) *cannot become positive,*

$$r'(t) \leqq 0 \text{ for } t_0 < t < \infty, \tag{14}$$

and, if t *is large enough,*

$$\int_t^{\infty} q(s)\,ds \leqq r(t) \leqq 1/(t - t_0) \qquad (t > t_0). \tag{15}$$

The inequalities (15) refine the consequences,

$$\int^{\infty} r^2(t)\,dt < \infty \text{ and } r(t) \to 0 \text{ as } t \to \infty, \tag{16}$$

of (10) and (12) (cf. [13], p. 375, where it is not assumed that $q \geqq 0$ but only that (10) holds and that the integral in (12) exists as an improper

integral; cf. also [4], p. 391). The first inequality in (15) is stated in [13], p. 377.

Proof of (*). Since (9) holds for large t, the assertion of (*) concerning (14) follows from (9) and $q \geqq 0$. It also follows from (9) that $r' + q \leqq 0$. Hence, a quadrature and the second of the relations (16) lead to the first of the inequalities (15). Finally, (9) implies that $r' + r^2 \leqq 0$. If $r(t) = 0$ for some large t, then $q \equiv 0$ and $r \equiv 0$, and so it can be supposed that $r(t) > 0$ for large t. Hence $r' + r^2 \leqq 0$ implies that $1 + r'/r^2 \leqq 0$, and so a quadrature gives $-r^{-1} + t \leqq$ const., which proves the second of the inequalities (15).

In order to obtain further results under the same assumption (10) as in (*), recourse must be had to the following theorem (cf. [2], p. 703, and for the following formulation, [6], p. 633): Even if just (10) (without (2_+)-(2_-) or $\operatorname{sgn} q = \text{const.}$) is assumed, (1) possesses a solution $y = y_0(t) \not\equiv 0$ which will be referred to as *the principal solution of* (1), is unique to a constant factor and is characterized by the property that

$$y_0(t) = o(|y(t)|) \text{ as } t \to \infty, \tag{17}$$

where $y(t)$ is any solution of (1) linearly independent of $y_0(t)$. An equivalent characterization of $y_0(t)$ is that, when the lower limit of integration is large enough (so large as to exceed the last zero of $y(t)$),

$$\int^{\infty} (y(t))^{-2} dt < \infty \text{ if and only if } y \not\equiv \text{const. } y_0, \tag{18}$$

where $y = y(t)$ denotes any solution ($\not\equiv 0$) of the non-oscillatory differential equation (1); cf. *loc. cit.* If $y = y(t) \not\equiv 0$ is not a principal solution, then

$$y_0(t) = y(t) \int_t^{\infty} (y(s))^{-2} ds$$

is defined for large t and is a principal solution.

(i) *Let* (1) *be non-oscillatory and let* $q(t)$ *be positive and monotone (so that* (11) *and* (2_-) *hold). Then the function* (3) *belonging to any solution of* (1) *is such as to satisfy* $h(t) \geqq 0$ *and* $dh(t) \leqq 0$, *which implies that* $h(\infty)$ *exists* $(< \infty)$ *and* $h(\infty) \geqq 0$. *There exists some solution* $y(t) \not\equiv 0$ *for which the sign of equality holds in* $h(\infty) \geqq 0$; *in fact, the principal solution* $y_0(t)$ *is such a* $y(t)$. *In order that* $h(\infty) = 0$ *holds for every solution* $y(t)$, *it is necessary and sufficient that*

$$\int^{\infty} tq(t)\,dt = \infty \tag{19}$$

(in the latter regard, cf. (12)-(13)).

Proof of (i). Since (7_), (11), (12) hold, only the assertions dealing with $h(\infty) = 0$ need to be proved. The first of these assertions is that

$$h_0(\infty) = 0, \tag{20}$$

where the subscript of h refers to the choice $y_0(t)$ of $y(t)$ in (3) (the corresponding notation, r_0, will be used in connection with (8)). Suppose that (20) is false. Then $h_0(t) > c$ holds for sufficiently large t and for some $c = \text{const.} > 0$. Since (3) and (8) imply that (9) can be written in the form

$$r' + h/y^2 = 0, \tag{21}$$

it follows that $-r_0' \geqq c/y_0^2$ (for large t), and so, by a quadrature,

$$-r_0(t) + \text{Const.} \geqq c \int^{t} (y_0(s))^{-2}\,ds.$$

But the expression on the right of this inequality cannot be bounded as $t \to \infty$, by (18), whereas the function on the left of the inequality is bounded, by (15). This contradication proves (20).

In order to prove the remaining part of (i), note that (19) fails to hold if and only if (1) has a solution $y(t) \not\equiv 0$ which is asymptotically proportional to t as $t \to \infty$ (for references, cf. [8]). On the other hand, if $y(t)$ is chosen to be positive for large t, then $y(t)$ is asymptotically proportional to t if and only if the (non-increasing) function $y'(t)$ has a *positive* limit $y'(\infty)$. Hence the statement of (i) concerning (19) is equivalent to the statement that $h(t) = y'^2 + qy^2 \geqq \text{Const.} > 0$ and $y > 0$ hold (for large t) if and only if $y'(t) \geqq \text{const.} > 0$ does (here $y(t)$ is a non-principal solution; cf. (18)).

Since the first part of the latter statement is clear, it is sufficient to show that if $h \geqq \text{Const.} > 0$, $y > 0$ hold for large t, then $y' \geqq \text{const.} > 0$ does. But $h(t) \geqq \text{Const.}$ and (21) imply that $-r' \geqq \text{Const.}/y^2$, hence a quadrature and (15), (18) show that

$$r(t) \geqq \text{Const.} \int_{t}^{\infty} (y(s))^{-2}\,ds.$$

If this inequality is multiplied by $y(t)$, then the expression on the right is

a principal solution $y_0(t)$ (cf. the formula line following (18)), and so the resulting inequality can be written in the form $y'(t) \geqq y_0(t)$. Since $y_0(t) > 0$ for large t, it follows, for reasons of convexity, that $y_0'(t) \geqq 0$ and $y_0(t) \geqq \text{const.} > 0$ for large t. Hence, $y'(t) \geqq \text{const.} > 0$ for large t. This completes the proof of (i).

3. *Negative $q(t)$.* While (*) and (i) have dealt with $q(t) > 0$, it will now be assumed that $q(t) < 0$; so that, if $q(t)$ is monotone, $-\infty \leqq q(\infty) \leqq 0$. If $-\infty < q(\infty) < 0$, the asymptotic behavior of every solution $y(t)$ and of its derivative is supplied by [12], pp. 60-61; cf. (39) below.

The assumption $q < 0$ alone (without (2_+)-(2_-)) implies that the principal solution $y_0(t)$ (to a constant factor) satisfies the inequalities

$$y_0(t) > 0, \qquad y_0'(t) < 0, \qquad y_0''(t) > 0 \qquad (0 \leqq t < \infty) \tag{22}$$

and that, for large t, every non-principal solution $y(t)$ (again to a constant factor) is subject to the inequalities

$$y(t) > 0, \qquad y(t) > 0, \qquad y''(t) > 0; \tag{23}$$

(Kneser; cf. [1], pp. 48-53, and, for other proofs, [2], p. 702, and [7]). There obviously exist solutions $y = y(t)$ satisfying (23) on the entire range $0 \leqq t < \infty$. Under the additional assumption that q is monotone (and that the above-mentioned case $-\infty < q(\infty) < 0$, which can be integrated by asymptotic formulae, is excluded), theorems (ii), (iii) below will describe the behavior of the functions (3) and (8).

(ii) *Let $q(t)$ be negative and monotone and $q(\infty) = 0$; so that (2_+) holds and*

$$0 > q \to 0 \text{ as } t \to \infty. \tag{24}$$

Then, for the principal solutions (22),

$$r_0(t) < 0, \qquad r_0'(t) > 0 \qquad (0 \leqq t < \infty) \tag{25}$$

and (what is more)

$$h_0(t) < 0, \qquad dh_0(t) \geqq 0 \qquad (0 \leqq t < \infty), \tag{26}$$

whereas for the non-principal solutions (23), *and if t is large enough,*

$$r(t) > 0, \qquad r'(t) < 0 \tag{27}$$

and (what is more)

$$h(t) > 0, \qquad dh(t) \geqq 0. \tag{28}$$

In addition, $r(\infty) = 0$ *holds for every solution and* $h_0(\infty) = 0$ *for the principal solution. Finally,* $h(\infty) = \infty$ *holds for the non-principal solutions* (23) *if and only if*

$$\int^{\infty} -tq(t)\,dt = \infty \qquad (q < 0). \tag{29}$$

The assertion $r(\infty) = 0$, made after (28), follows from Perron's refinement of Poincaré's theorem; cf. [9].

In contrast to (22) and (26), the inequalities (28) need not hold on $0 \leqq t < \infty$ (but only for large t) even if (23) holds for all $t \geqq 0$.

Proof of (ii). Since (9) is equivalent to (21), it is clear that (22) and (26) imply (25) and that (23) and (28) imply (29). Note that $dh \geqq 0$, by (2_+), (7_+).

Suppose, if possible, that $h_0(t) \geqq 0$ for some $t = t_0 \geqq 0$. Then $h_0(t) \geqq 0$ and, by (21), $r_0'(t) \leqq 0$ for $t \geqq t_0$. Since (22) implies that $r_0(t) < 0$, it follows that $r_0(t) \leqq -\text{const.} < 0$ for $t \geqq t_0$, which contradicts $r_0(\infty) = 0$. This proves (26), hence (25).

In order to prove (27) for large t, it is sufficient to prove that $h(t) > 0$ for some t. Since $r(t) > 0$ and $r(\infty) = 0$, it is clear that $r'(t) < 0$ for some t, say for $t = t_0$. Then (21) implies that $h(t_0) > 0$. This proves (28), hence (27).

Concerning the assertion $h_0(\infty) = 0$ of (ii), note that, by (21), $y_0(\infty)$ and $y_0'(\infty)$ exist (as finite limits). In addition, $y_0'(\infty) = 0$, since $y_0'(t)$ (< 0) has a finite integral over $0 \leqq t < \infty$. Hence $h_0(\infty) = 0$ follows from $y_0'(\infty) = 0$, $q(\infty) = 0$ and the boundedness of $y_0(t)$.

Finally, the proof of the assertion of (ii) concerning (29) and $h(\infty) = \infty$ is about the same as the proof of the assertion of (i) concerning (19) and $h(\infty) = 0$, and will therefore be omitted.

The following theorem (iii) describes the situation in the second of the cases

$$q(\infty) = 0, \qquad q(\infty) = -\infty, \qquad \text{where } q < 0. \tag{30}$$

in the same way as (ii) in the first of the cases (30).

(iii) *Let* $q(t)$ *be negative and monotone and* $q(\infty) = -\infty$; *so that* (2_-) *holds and*

$$0 > q(t) \to -\infty \text{ as } t \to \infty. \tag{31}$$

Then, for the principal solutions (22),

$$r_0(t) < 0, \qquad r_0'(t) < 0 \qquad (0 \leqq t < \infty) \tag{32}$$

and (what is more)

$$h_0(t) > 0, \qquad dh_0(t) \leqq 0 \qquad (0 \leqq t < \infty), \tag{33}$$

whereas for the non-principal solutions (23), *and if* t *is large enough,*

$$r(t) > 0, \qquad r'(t) > 0 \tag{34}$$

and (what is more)

$$h(t) < 0, \qquad dh(t) \leqq 0. \tag{35}$$

In addition,

$$r_0(\infty) = -\infty, h_0(\infty) = 0 \quad \text{or} \quad r(\infty) = \infty, h(\infty) = -\infty \tag{36}$$

according as the solution is principal or non-principal.

The first and third of the four assertions (36) are known; in fact, the classical proof (cf. [1], p. 54) for the case $-\infty < q(\infty) < 0$ is valid for the case $q(\infty) = -\infty$ also.

Proof of (iii). The proof of (33) (hence (32)) and of (35) (hence 34)) in (iii) is about the same as the proof of (26) (hence (25)) and of (28) (hence (27)) in (ii), and will therefore be omitted. What remain to be proved are the second and the fourth of the relations (36).

Since (22) implies that $\int^{\infty} -y_0'dt < \infty$, where $y_0' < 0$, and that y_0' is monotone, it is clear that $y_0'(\infty) = 0$. On the other hand, since $q < 0$, it is seen from (3) and (33) that $0 < h_0(t) < y_0'^2(t)$. Hence the second of the assertions (36) follows from $y_0'(\infty) = 0$.

Finally, since (23) implies that $y(\infty) = \infty$, hence $y^2(t) > \text{const.} > 0$ (for large t), the fourth of the assertions (36) follows from (2_+), (5) and (31).

The following theorem deals with the case of a monotone, negative $q(t)$ with $-\infty < q(\infty) < 0$, say

$$q(\infty) = -\omega^2, \quad \text{where} \quad 0 < \omega < \infty. \tag{37}$$

It is known that if $q(t) \to -\omega^2$ as $t \to \infty$, then the principal and non-principal solutions, $y = y_0(t)$ and $y = y(t)$, satisfy

$$r_0(t) \to -\omega \text{ and } r(t) \to \omega \text{ as } t \to \infty \tag{38}$$

(cf., e. g., [1], p. 54), and that (37) and the monotony of q (in fact, the boundedness of the variation of q on $0 \leqq t < \infty$) imply that

$$y_0(t) \sim \text{const.} \exp \int_0^t -(-q(s))^{\frac{1}{2}}ds, \quad y(t) \sim \text{Const.} \exp \int_0^t (-q(s))^{\frac{1}{2}}ds \tag{39}$$

as $t \to \infty$ (cf. [12], pp. 60-61).

(iv) *Let* $q(t)$ *be negative and monotone and satisfy* (37). *Then, according as* (2_+) *or* (2_-) *holds, the principal solutions* (22) *satisfy, for* $0 \leqq t < \infty$,

$$(40_+) \qquad r_0 < -\omega,\ r_0' > 0 \qquad \text{or} \qquad (40_-) \qquad r_0 > -\omega,\ r_0' < 0$$

and (*what is more*)

$$(41_+) \qquad h_0 < 0,\ dh_0 \geqq 0 \qquad \text{or} \qquad (41_-) \qquad h_0 > 0,\ dh_0 \leqq 0;$$

in addition, $h_0(\infty) = 0$. *On the other hand,*

$$(42) \qquad \pm \int^{\infty} \{\exp 2 \int^{t} (-q(s))^{\frac{1}{2}} ds\} dq(t) = \infty$$

is necessary and sufficient in order that the non-principal solutions satisfy

$$(43_+) \qquad h(\infty) = \infty \qquad \text{or} \qquad (43_-) \qquad h(\infty) = -\infty;$$

in which case, for large t,

$$(44_+) \qquad r > \omega,\ r' < 0 \qquad \text{or} \qquad (44_-) \qquad r < \omega,\ r' > 0$$

and (*what is more*)

$$(45_+) \qquad h > 0,\ dh \geqq 0 \qquad \text{or} \qquad (45_-) \qquad h < 0,\ dh \leqq 0.$$

If (42) *fails to hold, then* (43)-(45) *must be replaced by* (7) *and*

$$(46) \qquad h(\infty) \neq \pm \infty.$$

If (42) does not hold, then $h(\infty)$ can be made an arbitrary number $(\neq \pm \infty)$ by a suitable choice of the non-principal solution $y = y(t)$. For example, if $q(t) = -1$, so that (1) becomes $y'' - y = 0$, then, for the choice $y = e^t + ce^{-t}$ of the solution, the function (3) is $h(t) = -4c$ (and so $h(\infty) = -4c$).

Proof of (iv). The proofs of the assertions concerning r_0 and h_0 are similar to the proofs of the corresponding assertions in (ii) and will be omitted. The proof of the statements concerning (43) (hence (44) and (45)) and (46) follows from (5) and the last part of (39), leading to

$$(42 \text{ bis}) \qquad h(t) + \text{const.} \sim \text{Const.} \int^{t} \{\exp 2 \int^{s} (-q(u))^{\frac{1}{2}} du\} dq(s).$$

4. *The case where*

$$\text{(47)} \qquad \int^{\infty} q(t)\,dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right)$$

(*possibly just conditionally*). As mentioned after (15), the inequality $\int_t^{\infty} q(s)\,ds \leqq r(t)$ and the relations (16) do not depend on the assumption $q \geqq 0$ of (*) but only on (10) and (47).

In contrast to the inequality (15_1), which involves the remainder of (47), the assertions of (ii)-(iii), where $q < 0$ and (2_+)-(2_-) are assumed, lead to inequalities (again for r) which involve the square root of $-q$. In fact, if (2_+) and (24) hold, then, if t is large enough, $r_0 < 0$ and $r > 0$. Hence, by (26) and (28),

$$\text{(48)} \qquad 0 > r_0 > -q^* \quad \text{and} \quad r > q^*, \qquad \text{where } q^* = (-q)^{\frac{1}{2}} > 0.$$

Similarly, if (2_-) and (31) hold, then (33) and (35) imply that

$$\text{(49)} \qquad r_0 < -q^* \quad \text{and} \quad 0 < r < q^*, \qquad \text{where } q^* = (-q)^{\frac{1}{2}} > 0,$$

if t is large enough.

Since (48)-(49) are inequalities for the logarithmic derivative (8) of a principal or a non-principal solution, y_0 or y, quadratures lead to inequalities for y_0 and y themselves. There is a corresponding remark for (15_1), which holds for principal *and* non-principal solutions. In fact, a quadrature of (15_1), where (10) and (47) are assumed, shows that there belongs to every solution $y(t) \not\equiv 0$ of (1) a const. > 0 satisfying, for large t, the inequality

$$\text{(50)} \qquad y(t) > \text{const.} \exp \int_0^t Q(s)\,ds,$$

where

$$\text{(51)} \qquad Q(t) = \int_t^{\infty} q(s)\,ds$$

(cf. (47)). But the inequality (50) can be improved if the assumptions (10) and (47) are strengthened:

(†) *Let $q(t)$, where $0 \leqq t < \infty$, be any continuous function satisfying*

(47) *and having the property that, if* $Q(t)$ *is defined by* (51), *the differential equation*

$$x'' + 4Q^2(t)x = 0 \tag{52}$$

is non-oscillatory. Then

(α) *the differential equation* (1) *is non-oscillatory*;

(β) *corresponding to every solution* $x(t) \not\equiv 0$ *of* (52), *there exist two solutions of* (1), *say* $y = y_0(t)$ *and* $y = y^0(t)$, *satisfying*

$$0 < y_0(t) \leqq |x(t)|^{\frac{1}{2}} \exp \int_0^t Q(s)\,ds \leqq y^0(t) \tag{53}$$

for large t.

The assumptions that (47) holds and that (52) is non-oscillatory imply the inequalities (53) for certain solutions of (1). If the assumption that (52) is non-oscillatory is strengthened to

$$\int^{\infty} tQ^2(t)\,dt < \infty,$$

there result asymptotic formulae for the solutions of (1); cf. [18], theorem (i), p. 719.

Assertion (α) of (†) was proved in [3], p. 959. Assertion (β) of (†) will follow by a modification of the proof of (α).

Proof of (†). If an arbitrary solution $x(t) \not\equiv 0$ of (52) is replaced, when necessary, by $-x(t)$, then, since (52) is supposed to be non-oscillatory, it can be assumed that $x(t) > 0$ for large t. In terms of such an $x(t)$, define (for large t) a function by placing

$$z(t) = \{x(t)\}^{\frac{1}{2}} \exp \int_0^t Q(s)\,ds > 0. \tag{54}$$

Logarithmic differentiation of (54) gives

$$z'/z = \tfrac{1}{2}x'/x + Q, \tag{55}$$

whence another differentiation leads to

$$z'' - \{\tfrac{1}{2}p' - q + (\tfrac{1}{2}p + Q)^2\}z = 0, \tag{56}$$

where p denotes x'/x. In view of the latter definition, p belongs to (52) in the same way as r in (8) belongs to (1). Hence, what belongs to (52) in the

same way as (9) belongs to (1) is the Riccati equation $p' + p^2 + 4Q^2 = 0$.

It is seen that (56) can be written in the form

$$z'' + \{q + (\tfrac{1}{2}p - Q)^2\}z = 0, \qquad \text{where } p = x'/x. \tag{57}$$

Since the solution (54) of (57) does not vanish, the differential equation (57) is non-oscillatory. But the coefficient function q of (1) is majorized by the coefficient function $\{\cdots\}$ of (57). Hence (†) follows from the comparison theorem (iv) in [6], pp. 635-636, and its proof.

5. *A remark on the case* $q \geqq 0$. As mentioned at the beginning of Section **2**, condition (47) is necessary for (10) if $q \geqq 0$. Hence it is natural to ask whether or not the *converse* of assertion (α) of (†) is true in the case $q \geqq 0$, that is, whether or not (52) must be oscillatory if (1) is and if $q \geqq 0$. The answer to this question remains undecided even under the restriction that q tends monotonously to 0 as $t \to \infty$.

What is easy to see is that, even without the restriction just mentioned, the answer is in the affirmative if (52) is weakened to

$$x'' + Q^2 x = 0; \tag{58}$$

in addition, the assumption $q \geqq 0$ can be relaxed to $Q \geqq 0$, if (47) holds:

(I) *If* (1) *is non-oscillatory and* (47) *holds, and if the function* (51) *is non-negative, then* (58) *is non-oscillatory.*

In fact, if $y = y(t) \not\equiv 0$ is a solution of (1), then (16) is a consequence of (47), and so a quadrature of (9) leads to

$$-r + R + Q = 0, \text{ where } R(t) = \int_t^\infty r^2(s)\,ds$$

(for large t). This can be written in the form $\{R\} + 2RQ = 0$, if $\{R\}$ denotes $R' + R^2 + Q^2$. Consequently, $\{R\} \leqq 0$ (for large t), since $Q \geqq 0$ and $R \geqq 0$. Hence the statement concerning (58) follows from the criterion of [13], p. 368. In fact, the definition of $\{R\}$ shows that the limiting case, $\{R\} = 0$, of the inequality $\{R\} \leqq 0$, is the Riccati equation which belongs to (58) in the same way as (9) belongs to (1).

Under the assumptions of (I), assertion (β) of (†) has the following analogue:

(II) *If the conditions of* (I) *are satisfied, then, corresponding to every solution* $x = x(t) \not\equiv 0$ *of* (58), *there exists a solution* $y = y(t)$ *of* (1) *satisfying*

$$y(t) \geqq |x(t)| \exp \int_0^t Q(s)\,ds$$

for large t.

In particular, (1) has solutions satisfying

$$y(t) \geqq t \exp \int_0^t Q(s)\,ds$$

for large t, since $Q^2 \geqq 0$ implies that (58) has solutions satisfying $x(t) \geqq t$. If this remark is combined with the assumptions and the assertion of (I), it follows that (58) has solutions satisfying

$$x(t) \geqq t \exp \int_0^t \left(\int_s^\infty Q^2(u)\,du \right) ds.$$

Hence (1) has solutions for which

$$y(t) \geqq t \exp\{ \int_0^t Q(s)\,ds + \int_0^t \left(\int_s^\infty Q^2(u)\,du \right) ds \}$$

holds for large t.

The details of the proof of (II) will be omitted since it is similar to that of (β) in (†), if (54) is replaced by

$$z(t) = x(t) \exp \int_0^t Q(s)\,ds.$$

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] E. Borel, *Leçons sur les séries divergentes*, Paris (1901).

[2] P. Hartman, "Differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[3] ———, "On linear second order differential equations with small coefficients," *American Journal of Mathematics*, vol. 73 (1951), pp. 955-962.

[4] ———, "On non-oscillatory linear differential equations of second order," *ibid.*, vol. 74 (1952), pp. 107-126.

[5] ——— and A. Wintner, "On non-conservative linear oscillators of low frequency," *ibid.*, vol. 70 (1948), pp. 529-539.

[6] ——— and A. Wintner, "Oscillatory and non-oscillatory linear differential equations," *ibid.*, vol. 71 (1949), pp. 627-649.

[7] ——— and A. Wintner, "Linear differential and difference equations with monotone solutions," *ibid.*, vol. 75 (1953), pp. 731-743.

[8] ——— and A. Wintner, "On non-oscillatory linear differential equations," *ibid.*, vol. 75 (1953), pp. 717-730.

[9] O. Perron, "Ueber das infinitäre Verhalten der Integrale einer linearen Differentialgleichung zweiter Ordnung, wenn die charakterische Gleichung zwei gleiche Wurzeln hat," *Sitzungsberichte der Heidelberger Akademie*, 1917, Abhandlung 9.

[10] E. Picard, *Traité d'Analyse*, 3rd ed., vol. 3, Paris (1928).

[11] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *American Journal of Mathematics*, vol. 69 (1947), pp. 87-89.

[12] ———, "Asymptotic integrations of the adiabatic oscillator in its hyperbolic range," *Duke Mathematical Journal*, vol. 15 (1948), pp. 55-67.

[13] ———, "On the non-existence of conjugate points," *American Journal of Mathematics*, vol. 73 (1951), pp. 368-380.

DIVERGENCE OF VECTOR FIELDS AND DIFFERENTIAL EQUATIONS.*

By S. K. ZAREMBA.

The present note contains several propositions concerning the geometry of the integral curves of the differential equation

$$dx/X(x, y) = dy/Y(x, y). \tag{1}$$

All these propositions involve the divergence $X_x(x, y) + Y_y(x, y)$ of the vector (X, Y). The first part of this note gives a geometrically intrinsic form to the well-known formulae describing the variation of the integral curve of a differential equation, regarded as a function of the initial point. The second part deals with some immediate applications of the first part to the theories of limit cycles and first integrals; finally, in the third and fourth parts, we obtain a theorem which is partly reciprocal to Bendixson's criterion for the absence of closed integral curves.

I. The variation of the integral curve. It is well known [3] that if the functions $f(x, y)$ and $f_y(x, y)$ are continuous in a domain G, and if $\psi(x, x_1, \eta)$ is the solution of the differential equation $dy/dx = f(x, y)$ satisfying the initial conditions $x = x_1$, $y = \eta$, then

$$\psi_\eta(x_2, x_1, \eta) = \exp\{\int_{x_1}^{x_2} f_y[x, \psi(x, x_1, \eta)]\,dx\}, \tag{2}$$

provided that the arc of the curve $y = \psi(x, x_1, \eta)$ bounded by the points of abscissae x_1 and x_2 is contained in G.

In the more general case of an equation like (1), where neither of the inequalities $X(x, y) \neq 0$ and $Y(x, y) \neq 0$ holds throughout the domain under consideration, we cannot make use of this formula. However, the following proposition can then be applied:

THEOREM 1. *We assume that in a domain G the functions $X(x, y)$ and $Y(x, y)$ do not vanish simultaneously and have continuous first-order partial derivatives. Let (x_1, y_1) and (x_2, y_2) be any two points connected by an arc*

* Received June 12, 1953.

L of an integral curve $x = x(t)$, $y = y(t)$ *of* (1) *contained in G, the parameter t satisfying*

$$dx/X(x, y) = dy/Y(x, y) = dt \tag{3}$$

and $x(t_i) = x_i$, $y(t_i) = y_i$; $(i = 1, 2)$. *Finally, let* $\mathfrak{T}_i$ *be the orthogonal trajectory of* (1) *passing through the point* (x_i, y_i), *and let* u_i *be the parameter on* $\mathfrak{T}_i$ *satisfying*

$$-dx/Y(x, y) = dy/X(x, y) = du_i/\{[X(x, y)]^2 + [Y(x, y)]^2\} \tag{4}$$

and vanishing at the point (x_i, y_i); $i = 1, 2$.

Then, if we consider u_2 *as a function of* u_1 *defined by the condition that the values of* u_1 *and* u_2 *corresponding to each other should belong to points connected together by integral curves of* (1), *we have, for* $u_1 = 0$,

$$du_2/du_1 = \exp\{\int_{t_1}^{t_2} [X_x(x(t), y(t)) + Y_y(x(t), y(t))]\,dt\}. \tag{5}$$

Proof. L can be divided into a finite number of arcs, on each of which either $X(x, y)$ or $Y(x, y)$ is constantly different from zero, and it follows from the shape of (5) that it suffices to prove this formula for any of these arcs; for obvious reasons of symmetry, we can assume that, on the arc under consideration, $X(x, y) \neq 0$.

Let $\eta_1(u_1)$ and $\eta_2(u_1)$ be the ordinates of the intersections of the lines $x = x_1$ and $x = x_2$, in that order, with the integral curve of (1) passing through the point of $\mathfrak{T}_1$ which corresponds to the given value of u_1. By considering the curvilinear triangle formed by the aforesaid integral curve, by $\mathfrak{T}_1$ and by the line $x = x_1$, and by making this triangle tend to a point, we find, for $u_1 = 0$, $d\eta_1/du_1 = 1/X(x_1, y_1)$. Similarly, $du_2/d\eta_2 = X(x_2, y_2)$. Hence $du_2/du_1 = X(x_2, y_2)/X(x_1, y_1) \cdot d\eta_2/d\eta_1$. Substituting here, for $d\eta_2/d\eta_1$, the value given by (2) when $f(x, y) = Y(x, y)/X(x, y)$, we find

$$\frac{du_2}{du_1} = \frac{X(x_2, y_2)}{X(x_1, y_1)} \exp\{\int_{x_1}^{x_2} \frac{Y_y X - X_y Y}{X^2}\,dx\} = \exp\{\log \frac{X(x_2, y_2)}{X(x_1, y_1)} + \int_{t_1}^{t_2} \frac{Y_y X - X_y Y}{X}\,dt\},$$

which, after simplification, gives (5).

The interpretation of this formula in terms of two-dimensional flow is obvious, and (5) could also be obtained from the formula describing the variation of area in such a flow, but a rigorous proof on these lines seems to be longer than the preceding argument.

COROLLARY 2. *On the assumptions of Theorem* 1, *if, on* $\mathfrak{I}_1$ *and* $\mathfrak{I}_2$, *we introduce, instead of* u_1 *and* u_2, *parameters* v_1 *and* v_2 *satisfying*

$$-dx/Y(x,y) = dy/X(x,y) = dv_i \tag{6}$$

and vanishing at the point (x_i, y_i) $(i = 1, 2)$, *we find, for* $v_1 = 0$,

$$dv_2/dv_1 = [R(x_1, y_1)/R(x_2, y_2)]^2 \exp\{\int_{t_1}^{t_2} (X_x + Y_y)\,dt\}, \tag{7}$$

where $[R(x,y)]^2 = [X(x,y)]^2 + [Y(x,y)]^2$. *Similarly, on the same assumptions, if we introduce, on* $\mathfrak{I}_i$, *a parameter* σ_i *representing the length of the arc measured from the point* (x_i, y_i) *in the direction of the increasing* u_i*'s* $(i = 1, 2)$, *we find, for* $\sigma_1 = 0$,

$$d\sigma_2/d\sigma_1 = \{R(x_1, y_1)/R(x_2, y_2)\} \exp\{\int_{t_1}^{t_2} (X_x + Y_y)\,dt\}. \tag{8}$$

II. Miscellaneous applications of Theorem 1. In this section we propose to show how various propositions, which were obtained by means of apparently more complicated methods, can be represented as immediate, or almost immediate, consequences of Theorem 1.

We recall that a closed integral curve C of (1), which does not pass through singular points, is called a *limit cycle* if it is a limit set of integral curves passing through a certain neighborhood of C. The limit cycle C is called *stable* if it is the limit set, for $t \to +\infty$, of every integral curve passing through a certain neighborhod of C. It is called *unstable* if the same occurs for $t \to -\infty$ instead of $t \to +\infty$. Finally, a limit cycle is called *semi-stable* if the neighboring integral curves cituated on one side of it tend to it with $t \to +\infty$, and those situated on the other side tend to it with $t \to -\infty$. We have the following:

PROPOSITION 3. *A limit-cycle* C *of* (3) *is stable if*

$$\oint_C (X_x + Y_y)\,dt < 0, \tag{9}$$

and unstable if

$$\oint_C (X_x + Y_y)\,dt > 0. \tag{10}$$

Proof. Through an arbitrary point P of C, let us draw an orthogonal trajectory $\mathfrak{I}$ of (1), on which the parameter u is defined by (4) (u being substituted for u_i) and vanishes at P. Here, the formerly considered curves $\mathfrak{I}_1$ and $\mathfrak{I}_2$ coincide, and u_2 will now denote the value of the parameter u

belonging to the first point, say Q_2, of $\mathcal{T}$ met again after having started, from the point Q_1 of $\mathcal{T}$, defined by $u = u_1$, along an integral curve of (3) in the direction of the increasing t's.

According to Theorem 1, we have, for $u_1 = 0$,

$$du_2/du_1 = \exp\{\oint_C (X_x + Y_y)dt\},$$

which shows that, at this point, (9) implies $du_2/du_1 < 1$. Therefore, if (9) holds and u_1 is sufficiently small, $| u_2 | < | u_1 |$ *i. e.*, if Q_1 is sufficiently near to P, Q_2 falls between P and Q. From this, by a well-known type of argument [4], [5], [13], it follows that C is the limit set, for $t \to +\infty$, of the integral curve of (1) through Q_1. In other terms, (9) implies that C is stable. Similarly, the instability of C follows from (10).

The usual proof of this theorem is based on the theory of variational equations (e. g., [2]).

COROLLARY 4. *On a semi-stable limit cycle C of* (3), *we have*

$$\oint_C (X_x + Y_y)dt = 0.$$

(Cf. [7], where no proof is given.)

Hence, or directly from Proposition 3, we deduce the following:

COROLLARY 5. *A domain in which $X_x + X_y$ has a constant sign cannot contain semi-stable limit cycles.* (Cf. [7], where no proof is given.)

As is well known ([4], p. 78), if such a domain is simply connected, it cannot contain any closed integral curves, but otherwise stable—or unstable—limit cycles can occur.

PROPOSITION 6. *A domain in which $X_x + Y_y$ identically vanishes cannot contain limit cycles.*

Proof. If we revert to the notations of the proof of Theorem 1, we find that $du_2/du_1 = 1$. Indeed, since corresponding points on $\mathcal{T}_1$ and $\mathcal{T}_2$ are connected by integral curves of (1), a relationship of the type of (5) applies not merely when $u_1 = 0$, but whenever u_1 is sufficiently near to 0. Hence $u_2 = u_1$ identically, as $u_2 = 0$ when $u_1 = 0$. Now, if we assume the existence of a closed integral curve and make $\mathcal{T}_2$ coincide with $\mathcal{T}_1$, as in the proof of Proposition 3, this shows that all the neighboring integral curves are also closed.

This proposition is usually proved by an argument based on the invariance of areas under the group of the given differential equation (e. g., [5], pp. 92-93).

By a first integral of (1) valid in a set D of the (x, y)-plane, we understand a function, say $J(x, y)$, satisfying the following conditions: (i) $J(x, y)$ has continuous first-order partial derivatives throughout D; (ii) $J_x(x, y)$ and $J_y(x, y)$ do not vanish simultaneously in D except, possibly, at singular points of (1); (iii) $J(x, y)$ is constant along any arc of an integral curve of (1) contained in D.

The following theorem concerning first integrals and proved by H. H. Alden [1] can be obtained as an almost direct conseqquence of Corollary 2.

THEOREM 7. *Let the origin be a centre (in the sense of Poincaré)*[1] *of* (1) *and let D be a domain bounded by a closed integral curve and swept by closed integral curves of* (1) *surrounding the origin. Then if, throughout D, X and Y have continuous first order partial derivatives and do not vanish simultaneously, except at the origin, there exists a first integral of* (1) *valid in D.*

Proof. Let $J(x, y)$ be the area bounded by the integral curve, say C, of (1) passing through the point (x, y). If we make $J(0, 0) = 0$, J is evidently continuous throughout D and satisfies (iii). The coordinates of the running point on C being ξ and η, let $\xi = \phi(t)$, $\eta = \psi(t)$ be the solution of (3) satisfying $\phi(0) = x$, $\psi(0) = y$, and let ω be the smallest positive value of t for which $\phi(t) = x$, $\psi(t) = y$. Finally, let σ_1 and σ_2 be the lengths, counted as in Corollary 2, of the arcs on the integral curves of (4) passing through the points (x, y) and (ξ, η) respectively. It is easy to see that the partial derivative of J in the direction of the vector $(-Y, X)$ is equal to $\epsilon \oint_C d\sigma_2/d\sigma_1 \, ds$, s being the length of the arc of C and $\epsilon = \pm 1$, namely $\epsilon = +1$ if the vector $(-Y, X)$ points to the outside of C and $\epsilon = -1$ if it points to the inside. According to (8), this becomes equal to $R(x, y), z(x, y)$, where

$$(11) \qquad z(x, y) = \epsilon \int_0^{\omega} \exp\{\int_0^t [X_x(\phi(\tau), \psi(\tau)) + Y_y(\phi(\tau), \psi(\tau))] d\tau\} dt.$$

The expression obtained for the aforesaid partial derivative being continuous,

[1] We recall that, according to Poincaré [13], a singular point of (1) is called a *centre* if, at this point, $D(X, Y)/D(x, y) \neq 0$, and if all the integral curves passing through a certain neighborhood of this point are closed (Since Bendixson [4] this term has also been used in a much wider sense).

and the partial derivative of $J(x, y)$ in the direction of the vector (X, Y) identically vanishing, we find

$$(12) \qquad J_x(x, y) = -Y(x, y) \cdot z(x, y); \qquad J_y(x, y) = X(x, y) \cdot z(x, y).$$

In (11), ω is a continuous function of x and y, which is bounded in the neighborhood of the origin, since the absolute value of

$$d/dt \tan^{-1} y/x = (yX - xY)/(x^2 + y^2)$$
$$\simeq \{y[X_x(0, 0)x + X_y(0, 0)y] - x[Y_x(0, 0)x + Y_y(0, 0)y]\}/(x^2 + y^2)$$

has a positive lower bound in the neighborhood of the origin, the quadratic form in the numerator of the last expression being necessarily definite in the case of a centre [10], [13]. Hence $z(x, y)$ is bounded in the neighborhood of the origin and therefore $J_x(x, y)$ and $J_y(x, y)$ both tend to zero as $x, y \to 0$. On the other hand, according to (11) and (12), $|J_x(x, y)| + |J_y(x, y)| > 0$ whenever $|x| + |y| > 0$. Thus $J(x, y)$ is a first integral of (1) valid in D.

III. The existence of closed integral curves. As mentioned before, it is well known that if $X(x, y)$ and $Y(x, y)$ have continuous partial derivatives of the first order in a simply connected domain D, and if throughout D the divergence $\delta(x, y) \equiv X_x(x, y) + Y_y(x, y)$ has a constant sign, the differential equation (1) admits of no closed integral curve [2] contained in D.

Incidentally, this theorem, which was first stated by I. Bendixson in the form of a somewhat casual remark ([4], p. 78), seems now to be generally known owing to H. Dulac [7], who also made the rather obvious generalization of allowing $X(x, y)$ and $Y(x, y)$ to be multiplied by a variable factor $z(x, y)$.

We shall prove a theorem which can be regarded as partly reciprocal to Bendixson's theorem, and which shows that, in conjunction with a suitable factor $z(x, y)$, this criterion can give information about the existence of closed integral curves when it is most likely to be required.

Let us first examine the cases when the question of the existence of closed integral curves can be answered without deliberation. It is well known that the Kronecker index of a closed integral curve of (1) with respect to the vector (X, Y) is $+1$; cf. [9], [15], [16]. Hence the sum of the indices of the singular points in the interior region of a closed integral curve also amounts to $+1$, and therefore a simply connected domain containing no singular points with positive indices cannot contain closed integral curves.

---

[2] We recall, that according to the generally accepted convention, integral curves are considered to terminate at singular points. Exceptionally, however, we shall regard as a *closed integral curve* also the sum of a finite number of integral curves (in the preceding sense) if it forms a simple closed curve.

On the other hand, singular points of index greater than 1, and singular points of index 1, but of a topological type other than that of a centre, or of a nodal, or spiral point admit elliptic sectors in the sense of Brouwer [6], so that in their presence the problem of the existence of closed integral curves is solved in advance.

Apart from the case of a centre, in which this question is again automatically answered, the singular point must then be of the topological type common to nodal and spiral points, and characterized by the property that every integral curve passing through a certain domain U, containing the singular point, leads to this point in one, but always one, and always the same, of the two directions determined by the increasing, and by the decreasing t's. If, at the singular point, $D(X, Y)/D(x, y) \neq 0$, either all the integral curves reach it with definite tangents (and the point is then called *nodal*), or they do so winding themselves about it infinitely many times (and the point is then called a *spiral point*). In both cases,

$$D(X, Y)/D(x, y) > 0, \tag{12a}$$

since $I = +1$, and the singular point is reached only with $t \to \infty$ [10], [13].

The largest possible domain U, *i. e.* the locus of all the integral curves terminating at the singular point under consideration, will be called *the domain of influence* of the given singular point. The components of the boundary of this domain are clearly closed integral curves of (1), possibly passing through other singular points, or through the point at infinity, or even reduced to one of these points, which must in any case be excluded from U. Thus the domain of influence can also be described as the largest domain containing the given singular point, but containing neither parts of closed integral curves of (1), nor other singular points, or points at infinity. With a few additional, but not very restrictive assumptions listed below, we can now state the following theorem:

THEOREM 8. *Let the functions $X(x, y)$ and $Y(x, y)$ have continuous first and second order partial derivatives throughout the (x, y)-plane, and let Equation* (1) *have a spiral, or nodal, point with a simply connected domain of influence, which will be denoted by D. If at this singular point $\delta(x, y) \equiv X_x + Y_y \neq 0$, and if $F(x, y)$ is a function not vanishing at the singular point and admitting continuous first order partial derivatives throughout D, but otherwise arbitrarily given, there exists a function $z(x, y)$ defined throughout D with continuous first order partial derivatives and identically satisfying the condition*

$$[z(x, y) \cdot X(x, y)]_x + [z(x, y) \cdot Y(x, y)]_y = F(x, y). \tag{13}$$

Moreover, if $X(x, y)$, $Y(x, y)$ and $F(x, y)$ are analytic in D, $z(x, y)$ is also analytic in D; on the other hand, if $F(x, y)$ has a constant sign in D, $z(x, y)$ also has a constant sign in D.

Proof. Obviously, we can assume that the singular point under consideration is at the origin of the coordinates. Without impairing the generality of our argument, we can also assume that

$$\delta(0, 0) < 0. \tag{14}$$

Equation (13) can be rewritten in the following form,

$$X(x, y)z_x + Y(x, y)z_y = F(x, y) - \delta(x, y) \cdot z, \tag{15}$$

which shows that this is a first order linear partial differential equation in z, whose characteristics are given by the equations

$$dx/X(x, y) = dy/Y(x, y) = dz/\{F(x, y) - z \cdot \delta(x, y)\}. \tag{16}$$

The projections of these characteristics on the (x, y)-plane are clearly the integral curves of (1). If $x = x(t)$, $y = y(t)$ is a solution of (3), the unknown function z, considered along the corresponding integral curve, as a function of t [where t is determined by (3)], satisfies the ordinary linear differential equation

$$dz/dt + \delta[x(t), y(t)] \cdot z = F[x(t), y(t)].$$

Because of (14), the point $x(t), y(t)$ tends to the origin as $t \to +\infty$ [10], [13]. Hence our solution of this equation must remain finite for $t \to +\infty$. But it is easy to see that there is only one such solution; it can be expressed by the formula

$$z[x(t), y(t)] = -\int_t^{+\infty} F[x(T), y(T)] \exp\{\int_t^T \delta[x(\tau), y(\tau)]d\tau\}dT \tag{17}$$

and we find that

$$\lim_{t \to +\infty} z[(t), y(t)] = F(0, 0)/\delta(0, 0). \tag{18}$$

On the other hand, it is fairly obvious that, in our case, there exists a family of homothetic ellipses centred at the origin and, in a certain neighborhood of the latter, nowhere tangent to the vector (X, Y). It is clearly sufficient to prove it for the linearized vector field (X_0, Y_0) defined by $X_0 = Ax + By$, $Y_0 = Cx + Dy$, where

$$A = X_x(0, 0), \qquad B = X_y(0, 0), \qquad C = Y_x(0, 0), \qquad D = Y_y(0, 0),$$

and this can best be done as follows:

In view of $A + D = \delta(0, 0) \neq 0$, we can assume $AD > 0$, for, if this inequality did not hold, we could, by a rotation of the axes of co-ordinates (which leaves $A + D$ invariant), make $A = D$ and hence $AD > 0$. Now the equation of the family of ellipses can be written as $ax^2 + by^2 = c$, where c is the parameter of the family. The condition that the ellipses should never be tangent to the vector (X_0, Y_0) is that the quadratic form $a(Ax + By)x + b(Cx + Dy)y$ should be definite; verifying that this can be obtained with positive values of a and b is a matter of simple algebra and is based on the inequality $AD(AD - BC) > 0$.

Any of the aforenamed ellipses, if sufficiently small, is intersected exactly once by every integral curve of (1) passing through a certain neighborhood of the origin. Having arbitrarily chosen one of these ellipses and denoting by w the parameter of a point thereon, let us, further, denote by $[x(t, w), y(t, w)]$ the solution of (3) determined by the condition that $[x(0, w), y(0, w)]$ should be the point of the ellipse corresponding to the given w. We have clearly

$$D(x, y)/D(t, w) \neq 0 \tag{19}$$

throughout D, with the obvious exception of the origin. Then (17), with the substitution of $x(t, w)$ and $y(t, w)$ for $x(t)$ and $y(t)$, defines $z(x, y)$ throughout D, and, since the relevant integrals are uniformly convergent, this is a function which has continuous first order partial derivatives in x and y and satisfies (15); further, the shape of (17) shows that if $F(x, y)$ has a constant sign in D, $z(x, y)$ also has a constant sign.

The only point of D where this argument does not apply, and indeed where $z(x, y)$ has not yet been defined, is the origin. Owing to (18), if we make $z(0, 0) = F(0, 0)/\delta(0, 0)$, $z(x, y)$ becomes continuous at the origin, at least when the origin is approached along an integral curve of (1).

Now we shall see that if $X(x, y)$, $Y(x, y)$ and $F(x, y)$ are analytic throughout D, $z(x, y)$ is also analytic throughout D. The point $x = y = 0$, $z = F(0, 0)/\delta(0, 0)$ is a singular point of (16), at which the Jacobian of the denominators equals $-\delta(0, 0)D(X, Y)/D(x, y) \neq 0$.

Poincaré's equation in S [14] becomes

$$\begin{vmatrix} X_x - S & X_y & 0 \\ Y_x & Y_y - S & 0 \\ F_x - z\delta_x & F_y - z\delta_y & -\delta - S \end{vmatrix} = 0.$$

It has one positive root, $S = -\delta(0, 0)$, and two other roots which are both negative in the case of a nodal point of (1) and imaginary conjugate with a negative real part in the case of a spiral point of (1). In both cases [14,

p. 170 and p. 175 in the *Oeuvres*], there exists exactly one solution of (15) passing through the singular point of (16), and this solution is analytic. It clearly must coincide with $z(x, y)$, as defined by (17); therefore $z(x, y)$ is analytic throughout D.

IV. The non-analytic case. In the non-analytic case, no indication about the behavior of $z_x(x, y)$ and $z_y(x, y)$ at the origin can be derived from the literature concerning singular points of systems of differential equations (*e. g.*, [9], [11], [12]); a study *ad hoc* is, however, possible owing to Theorem 1.

We first reduce the general case to that when

$$\delta(0, 0) = F(0, 0);\ \delta_x(0, 0) = F_x(0, 0);\ \delta_y(0, 0) = F_y(0, 0). \tag{20}$$

It is easily verfiable that, in view of (12a), this can be achieved by substituting, in (15), $(Kx + Ly + M) \cdot X(x, y)$ and $(Kx + Ly + M) \cdot Y(x, y)$ respectively for X and Y, the constants K, L, M being determined by a set of three linear equations. Then the originally required factor is $(Kx + Ly + M) \cdot z(x, y)$. The continuity of $z_x(x, y)$ and $z_y(x, y)$ at the origin will now follow from the relations

$$\lim z_x(x, y) = 0, \qquad \lim z_y(x, y) = 0; \qquad (x, y) \to (0, 0).$$

which, in turn, are equivalent to the statement that the partial derivatives of $z(x, y)$ in the directions of the two mutually orthogonal vectors (X, Y) and $(-Y, X)$ tend to zero as the point (x, y) tends to the origin.

On the other hand, the expression given for $z(x, y)$ by (17) is linear in $F(x, y)$ and is constantly equal to 1 if $F(x, y) \equiv \delta(x, y)$; hence

$$z - 1 = -\int_t^{+\infty} G[x(T), y(T)] \exp\{\int_t^T \delta[x(\tau), y(\tau)]d\tau\}dT; \tag{21}$$

where $G(x, y) \equiv F(x, y) - \delta(x, y)$.

Thus

$$dz/dt = G[x(t), y(t)] + \delta[x(t), y(t)] \int_t^{+\infty} G[x(t), y(t)] \times \exp\{\int_t^T \delta[x(\tau), y(\tau)]d\tau\}dT.$$

As it was pointed out, the neighborhood of a nodal, or spiral, point can be covered by a family of homothetic ellipses without contact with the integral

curves. Hence there exists a positive number α such that, on any integral curve of (1) in a neighborhood of the origin,

$$[x(T)]^2 + [y(T)]^2 \leqq \alpha^2\{[x(t)]^2 + [y(t)]^2\} \text{ whenever } t \leqq T.$$

Owing to (14) and (20), there exist numbers β, δ_1, and δ_2 such that, in a neighborhood of the origin

$$(22) \qquad |G(x, y)| < \beta \cdot (x^2 + y^2); \qquad (23) \qquad \delta_1 < \delta(x, y) < \delta_2 < 0.$$

Hence, in a sufficient small neighborhood of the origin,

$$(23a) \quad |dz/dt| < \{[x(t)]^2 + [y(t)]^2\}\{\beta - \alpha^2\beta\delta_1 \int_t^{+\infty} e^{\delta_2(T-t)} dT\}$$
$$= \beta[1 + \alpha^2\delta_1/\delta_2]\{[x(t)]^2 + [y(t)]^2\}.$$

The partial derivative of z along the vector (X, Y) equals $R^{-1}dz/dt$ and since, in the neighborhood of the origin,

$$(24) \qquad [R(x, y)]^2 \geqq \lambda^2(x^2 + y^2),$$

where the constant λ is positive, it follows from (23a) that the partial derivative of z along the vector (X, Y) tends to zero as $(x, y) \to (0, 0)$.

In order to obtain an evaluation of the partial derivative of z along the vector $(-Y, X)$, we assume that a point (x_0, y_0) of D, sufficiently near to the origin, is given. Through this point, we trace a solution,

$$(25) \qquad x = \xi(u_1), \qquad y = \eta(u_1),$$

of (4), the parameter u_1 vanishing at the point (x_0, y_0). The solution of (3) satisfying (25) for $t = 0$ can be written as

$$(26) \qquad x = x(t, u_1), \qquad y = y(t, u_1).$$

Accordingly, we rewrite (21) as follows:

$$(27) \qquad z[\xi(u_1), \eta(u_1)] = 1 - \int_0^{+\infty} G[x(t, u_1), y(t, u_1)]$$
$$\times \exp\{\int_0^t \delta[x(\theta, u_1), y(\theta, u_1)]\, d\theta\} dt.$$

In order to facilitate the differentiation of this expression with respect to u_1, we will change the variable of integration in the following manner: If u_1 is sufficiently near to zero, every point of (26) corresponding to a positive t is connected by an integral curve of (4) with a point $x = x(T, 0)$, $y = y(T, 0)$

of the curve (26) corresponding to $u_1 = 0$, so that t can be regarded as a function, say $\chi(T, u_1)$ of T and u_1. Let

$$\phi(T, u_1) = x[\chi(T, u_1), u_1], \qquad \psi(T, u_1) = y[\chi(T, u_1), u_1].$$

Noting that, for a constant T, $x = \phi(T, u_1)$, $y = \psi(T, u_1)$ are parametric equations of an integral curve of (4), we can introduce on it, for a while, a parameter u_2 satisfying (4). Then, applying Formula (5), we find, for $u_1 = 0$,

$$\partial\phi/\partial u_1 = -Y(x, y)/[R(x, y)]^2 \exp\{\int_0^T \delta[\phi(\tau, 0), \psi(\tau, 0)]d\tau\},$$

$$\partial\psi/\partial u_1 = X(x, y)/[R(x, y)]^2 \exp\{\int_0^T \delta[\phi(\tau, 0), \psi(\tau, 0)]d\tau\},$$

where $x = \phi(T, 0)$, $y = \psi(T, 0)$, and therefore, owing to (24),

$$|\phi_{u_1}(T, 0)|, |\psi_{u_1}(T, 0)| < \lambda^{-1}\{[\phi(T, 0)]^2 + [\psi(T, 0)]^2\}^{-\frac{1}{2}}. \tag{28}$$

With the new variable of integration, (27) becomes

$$z[\xi(u_1), \eta(u_1)] = 1 - \int_0^{+\infty} G[\phi(T, u_1), \psi(T, u_1)] \times \exp\{I[\chi(T, u_1), u_1]\}\chi_T(T, u_1)dT, \tag{29}$$

where $I(t, u_1) = \int_0^t \delta[x(\theta, u_1), y(\theta, u_1)]d\theta$, and consequently

$$I[\chi(T, u_1), u_1] = \int_0^T \delta[\phi(\Theta, u_1), \psi(\Theta, u_1)]\chi_T(\Theta, u_1)d\Theta. \tag{30}$$

Of course,

$$\chi_T(T, 0) = 1, \tag{31}$$

but in order to differentiate (29) and (30), we need an explicit expression for χ_T. It can be obtained from (7), where we have to substitute $-Y$ and X for X and Y respectively. Denoting by v a parameter satisfying (6), we find

$$\chi_T(T, u_1) = \frac{\{R[\phi(T, 0), \psi(T, 0)]\}^2}{\{R[\phi(T, u_1), \psi(T, u_1)]\}^2} \exp\{\int_{v_1}^{v_2} [X_y - Y_x]dv\}, \tag{32}$$

where in X_y and Y_x, the values of x and y as functions of v should be substituted, and where v_1 and v_2 denote the values of v at the intersections with the curves $x = \phi(T, 0)$, $y = \psi(T, 0)$ and $x = \phi(T, u_1)$, $y = \psi(T, u_1)$ respectively.

If, in the neighborhood under consideration, $|X_x|$, $|X_y|$, $|Y_x|$, $|Y_y| < \omega$, the partial derivative, with respect to u_1, of the first factor in (32) is, owing

to (28), less than $8\omega\lambda^{-2}\{[\phi(T,0)]^2+[\psi(T,0)]^2\}^{-1}$ in absolute value for $u_1=0$. As to the integral in the second factor, by relating v to a parameter u_2 satisfying (4), and by applying (5), we obtain

$$\lim_{u_1\to 0}(v_2-v_1)/u_1=\{R[\phi(T,0),\psi(T,0)]\}^{-2}\times\exp\{\int_0^T\delta[\phi(\Theta,0),\psi(\Theta,0)]d\Theta\},$$

and therefore, for $u_1=0$,

$$d/du_1\int_{v_1}^{v_2}(X_y-Y_x)dv=\frac{X_y[\phi(T,0),\psi(T,0)]-Y_x[\phi(T,0),\psi(T,0)]}{\{R[\phi(T,0),\psi(T,0)]\}^2}\times\exp\{\int_0^T\delta[\phi(\Theta,0),\psi(\Theta,0)]d\Theta\},$$

which is less than $2\omega\lambda^{-2}\{[\phi(T,0)]^2+[\psi(T,0)]^2\}^{-1}$ in absolute value. Both factors in (32) becoming equal to 1 when $u_1=0$, we conclude that

$$|\chi_{T,u_1}(T,0)|<10\omega\lambda^{-2}\{[\phi(T,0)]^2+[\psi(T,0)]^2\}^{-1}. \tag{33}$$

It follows from our assumption that there exists a number ρ such that, in the neighborhood under consideration, $|G_x(x,y)|$, $|G_y(x,y)|<\rho(x^2+y^2)^{\frac{1}{2}}$, Then, according to (28),

$$|\partial G[\phi(T,u_1),\psi(T,u_1)]/\partial u_1|<2\rho/\lambda. \tag{34}$$

Similarly, if $|\delta_x(x,y)|$, $|\delta_y(x,y)|<\sigma$, we have, for $u_1=0$,

$$|\partial\delta[\phi(T,u_1),\psi(T,u_1)]/\partial u_1|<2\sigma\{[\phi(T,0)]^2+[\psi(T,0)]^2\}^{-\frac{1}{2}}/\lambda. \tag{35}$$

If we differentiate (30), with respect to u_1 and take into account (23), (31), (33), and (35), we find, for $u_1=0$,

$$|\partial I[\chi(T,u_1),u_1]/\partial u_1|<2\sigma/\lambda\int_0^T\{[\phi(\Theta,0)]^2+[\psi(\Theta,0)]^2\}^{-\frac{1}{2}}d\Theta+10\omega|\delta_1|/\lambda^2\int_0^T\{[\phi(\Theta,0)]^2+[\psi(\Theta,0)]^2\}^{-1}d\Theta.$$

On the other hand, from the properties of the previously introduced constant α it follows that, for $0\leqq\Theta\leqq T$,

$$[\phi(\Theta,0)]^2+[\psi(\Theta,0)]^2\geqq\alpha^{-2}\{[\phi(T,0)]^2+[\psi(T,0)]^2\}.$$

Hence there exists a constant ν satisfying, for $u_1=0$,

$$|\partial I[\chi(T,u_1),u_1]/\partial u_1|<\nu T\cdot\{[\phi(T,0)]^2+[\psi(T,0)]^2\}^{-1}. \tag{36}$$

Now, if we differentiate (29) with respect to u_1, noting that the relevant integrals are uniformly convergent and taking into account (22), (23), (31), (33), (34), and (36), we find that $| \partial z[\xi(u_1), \eta(u_1)]/\partial u_1 |$ is less than

$$\int_0^{+\infty} [(2\rho\lambda + 10\omega\beta)/\lambda^2 + \beta\nu T] e^{\delta_2 T} dT = (2\rho\lambda + 10\omega\beta)/(|\delta_2| \lambda^2) + \beta\nu/\delta_2^2,$$

hence bounded in the neighborhood of the origin, and since, when multiplied by $R(x_0, y_0)$, it gives the partial derivative of $z(x, y)$ in the direction of the vector $(-Y, X)$, the latter derivative tends to zero as $(x, y) \to (0, 0)$.

Thus both partial derivatives of $z(x, y)$, which are continuous in D outside the origin, tend to 0 as $x, y \to 0$. In view of the fact that $z(x, y)$ is continuous at the origin when approached along integral curves of (1), now it is easy to see that this function is continuous with continuous partial derivatives throughout D. Hence the proof of Theorem 8 is complete.

BOULTON PAUL AIRCRAFT LIMITED,
WOLVERHAMPTON, ENGLAND.

REFERENCES.

[1] H. H. Alden, "Solution of $f(x, y)\partial z/\partial x + g(x, y)\partial z/\partial y = 0$ in a neighborhood of a singular point," *American Journal of Mathematics*, vol. 56 (1934), pp. 593-612.

[2] A. Andronow, "Les cycles limites de Poincaré et la théorie des oscillations auto-entretenues," *Comptes Rendus de l'Académie des Sciences*, vol. 189 (1929), pp. 559-561.

[3] I. Bendixson, "Démonstration de l'existence de l'intégrale d'une équation aux dérivées partielles linéaire," *Bulletin Société Mathématique de France*, vol. 24 (1896), pp. 220-225.

[4] ———, "Sur les courbes définies par des équations différentielles," *Acta Mathematica*, vol. 24 (1900), pp. 1-88.

[5] L. Bieberbach, *Theorie der Differentialgleichungen*, 3rd ed., Berlin, 1930.

[6] L. E. J. Brouwer, "On continuous vector distributions on surfaces," *Proceedings Koninklijke Akademie van Wetenshapen*, vol. XII_2 (1910), pp. 716-734.

[7] H. Dulac, "Recherche des cycles limites," *Comptes Rendus de l'Académie dees Sciences*, vol. 204 (1937), pp. 1703-1706.

[8] J. Haag, "Cols, noeuds et foyers," *Bulletin des Sciences Mathématiques*, ser. 2, vol. 74 (1950), pp. 167-192.

[9] B. von Kerékjártó, "On the variation of the tangent to a simple closed curve," *Proceedings of the London Mathematical Society*, ser. 2, vol. 23 (1925), pp. XXXIX-XL.

[10] O. Perron, "Über die Gestalt der Integralkurven einer Differentialgleichung erster Ordnung in der Umgebung eines singulären Punktes," *Mathematische Zeitschrift*, vol. 15 (1922), pp. 121-146 and vol. 16 (1923), pp. 273-295.

[11] ———, "Über Stabilität und asymptotisches Verhalten von Differentialgleichungssystem," *ibid.*, vol. 29 (1929), pp. 129-160.

[12] I. Petrowsky, "Über des Verhalten der Integralkurven eines Systems gewöhnlicher Differentialgleichungen in der Nähe eines singulären Punktes," *Recueil Mathematique*, vol. 41 (1934), pp. 107-155.

[13] H. Poincaré, "Mémoire sur les courbes définies par une équation différentielle," *Journal de Mathématiques Pures et Appliquées*, ser. 3, vol. 7 (1881), pp. 375-422 and ser. 3, col. 8 (1882), pp. 251-296; or *Oeuvres*, vol. I.

[14] ———, "Sur les courbes définies par les équations différentielles," *ibid.*, ser. 4, vol. 1 (1885), pp. 167-244 and ser. 4, vol. 2 (1886), pp. 151-217; or *Oeuvres*, vol. I.

[15] G. N. Watson, "A problem of analysis situs," *Proceedings of the London Mathematical Society*, ser. 2, vol. 15 (1916), pp. 227-242.

[16] S. K. Zaremba, "Sur la variation de la tangente à une courbe fermée simple de Jordan," *Annales de la Société Polonaise de Mathématique*, vol. 12 (1933), pp. 55-56.

ON THE HOMOTOPY TYPE OF AN A_n^3-POLYHEDRON ($n \geqq 3$).*

By KENICHI SHIRAIWA.

1. Introduction. It is an important topological problem to determine the homotopy type of a space by means of its algebraically computable invariants. Only certain special cases have been solved.

In 1949, J. H. C. Whitehead [3] proved that two simply-connected 4-demensional polyhedra are of the same homotopy type if and only if their cohomology systems are properly isomorphic. He then introduced a special type of a polyhedron, a generalization of a simply connected 4-dimensional polyhedron named an A_n^2-complex which is $(n-1)$-connected and at most $(n+2)$-dimensional. He proved an analogous result in [4]. A related result is the proof of S. C. Chang [6] that any A_n^2-complex ($n \geqq 3$) has the same homotopy type as a union of the elementary polyhedra (See section 4 below) with a single point in common.

This raises the problem of the homotopy type of an A_n^3-polyhedron, which is $(n-1)$-connected and at most $(n+3)$-dimensional. May an analogous result be obtained? H. Uehara has established [10] the relevant fact that, roughly speaking, the homotopy type of an A_n^3-complex ($n > 3$) with vanishing $(n+1)$-homotopy group is completely determined by means of its cohomology groups and Sq^2 of Steenrod [8].

The main purpose of this paper is to determine, in general, the homotopy type of an A_n^3-complex ($n \geqq 3$) by means of certain cohomology operations. These operation are Steenrod square, Sq^2, Adem's operation [9], and its modification introduced by N. Shimada-H. Uehara [11]. Our method and the theorem obtained here are quite analogous to those of J. H. C. Whitehead [3], [4]. The author is deeply grateful to Messrs. N. Simada and H. Uehara for their kind criticisms and encouragement.

Throughout this paper, we assume $n \geqq 3$.

2. A_n^3-cohomology systems. First we shall consider the case $n > 3$. Let H^r ($r = 0, 1, 2, \cdots$) be finitely generated abelian groups which satisfy the following conditions.

* Received May 8, 1953.

i) $H^r = 0$ for all values of $r \neq 0, n, n+1, n+2, n+3$.

ii) H^0 is a free cyclic group.

iii) H^n is a free abelian group.

We now introduce another group, using the notation that if A is any additive group, then ${}_mA$ is the subgroup of all elements $a \in A$ such that $ma = 0$, and A_m is the residue group $A - Am$ $(m > 0)$. We define $\Delta^*{}_m({}_mH^{r+1})$ $(m = 2, 4)$, to be the image of ${}_mH^{r+1}$ in an isomorphism, $\Delta^*{}_m$, and define $H^r(m)$ as $H^r(m) = H^r{}_m + \Delta^*{}_m({}_mH^{r+1})$. We do not here regard $\Delta^*{}_m$ as uniquely determined by the system, but we do postulate a uniquely determined homomorhpism $\Delta_m : H^r(m) \to H^{r+1}$ such that $\Delta^{-1}{}_m(0) = H^r{}_m$ and $\Delta_m \Delta^*{}_m = 1$. We also postulate $\mu_{m,0} : H^r \to H^r(m)$ as the natural map onto the summand $H^r{}_m \subset H^r(m)$. We next define $\mu_{2,4} : H^r(4) \to H^r(2)$ as follows: $\mu_{2,4} \mid H^r{}_4$ is the natural map onto the summand $H^r{}_2 \subset H^r(2)$ and $\mu_{2,4} \mid \Delta^*{}_4({}_4H^{r+1})$ is defined by the formula

$$\mu_{2,4}(x) = \Delta^*{}_2(2\Delta_4(x)) \quad \text{for } x \in \Delta^*{}_4({}_4H^{r+1}).$$

We also define $\mu_{4,2} : H^{n+3}(2) = H^{n+3}{}_2 \to H^{n+3}{}_4 = H^{n+3}(4)$ by

$$\mu_{4,2}(x \bmod 2H^{n+3}) = 2x \bmod 4H^{n+3} \quad \text{for } x \in H^{n+3}.$$

Finally, we postulate $\gamma_r : H^r(2) \to H^{r+2}(2)$, $(r = n, n+1)$, and

$$\Phi_1 : N_1 = (\gamma_n \mu_{2,4})^{-1}(0) \to H^{n+3}(2)/\gamma_{n+1}H^{n+1}(2) = M_1,$$

$$\Phi_2 : N_2 = \gamma^{-1}{}_n(0) \cap (\gamma_{n+1}\Delta_2)^{-1}(0) \cap N \to 2H^{n+3}(4)/\mu_{4,2}\gamma_{n+1}H^{n+1}(2) = M_2$$

where N is the subgroup of $H^n(2)$ consisting of all elements v such that $\Delta_2 v = 0$. They may be arbitrary homomorphisms. By an $A_n{}^3$-cohomology system H $(n > 3)$ we shall mean a set of groups H^r $(r = 0, 1, \cdots)$ $H^r(2)$ $(r = n, n+1, n+2, n+3)$ $H^r(4)$ $(r = n, n+3)$ together with homomorphisms Δ_m, $\mu_{m,0}$ $(m = 2, 4)$, γ_r $(r = n, n+1)$, Φ_i $(i = 1, 2)$ of the kind just described.

Let H, $H^\dagger$ be two $A_n{}^3$-cohomology systems $(n > 3)$. By a (Δ, μ)-homomorphism $f : H \to H^\dagger$, we mean a map such that $f \mid H^r$, $f \mid H^r(m)$ are homomorphisms into the corresponding groups $H^{\dagger r}$, $H^{\dagger r}(m)$ and $f\Delta_m = \Delta_m f$, $f\mu_{m,0} = \mu_{m,0} f$ where Δ_m, $\mu_{m,0}$ mean the same in $H^\dagger$ as in H. If f is a (Δ, μ)-homomorphism and $f\gamma_r = \gamma_r f$, then $f(N_1) \subset N_1{}^\dagger$, $f(N_2) \subset N_2{}^\dagger$ and f induces the homomorphisms M_1 into $M_1{}^\dagger$ and M_2 into $M_2{}^\dagger$. Therefore, when f commutes with γ_r, $f\Phi_i$, $\Phi_i f$ have the same meaning. By a proper homomorphism $f : H \to H^\dagger$, we mean a (Δ, μ)-homomorphism $f : H \to H^\dagger$ which commutes with γ_r and also Φ_i.

We have defined an $A_n{}^3$-cohomology system when $n > 3$. We shall modify these definitions in case $n = 3$. This happens because in an $A_3{}^3$-complex its cohomology ring has a non-trivial multiplication.

Let R be a ring which satisfies the following conditions (Cf. J. H. C. Whitehead [3]):

i) $R = \sum H^r(m)$, $\quad (r = 0, 1, \cdots)$ $(m = 0, 2, 3, \cdots)$.

ii) All $H^r(m)$ are finitely generated abelian groups which form a spectrum of cohomology groups in the sense of J. H. C. Whitehead [2] with homomorphisms $\Delta_q : H^r(q) \to H^{r+1}$, $\mu_{p,q} : R^r(q) \to H^r(p)$.

iii) $H^r(m) = 0$ if $r \neq 0, 3, 4, 5, 6$.

iv) H^0 is a free cyclic group whose generator is a unit element of R.

v) In R the multiplication satisfies the following conditions:

a) if $x \in H^p(m)$, $y \in H^q(m)$, then $xy \in H^{p+q}(m)$ and $xy = (-1)^{pq} yx$,

b) $\mu_{m,r}(x)\mu_{m,s}(y) = [m(m, r, s)/(m, r)(m, s)]\mu_{m,c}(xy)$ where $x \in H^p(r)$, $y \in H^q(s)$, $c = (r, s)$, $m > 0$.

If, in the definition of an $A_n{}^3$-cohomology system $(n > 3)$, we replace the groups H^r, $H^r(m)$ by R and homomorphisms Δm, $\mu_{m,0}$ $(m = 2, 4)$ by Δ_p, $\mu_{p,q}$, then we get a definition of an $A_3{}^3$-cohomology system. By a (Δ, μ)-homomorphism between two $A_3{}^3$-cohomology systems, we shall mean a ring homomorphism of one cohomology system into the other which commutes with Δ_p and $\mu_{p,q}$. A proper homomorphism is the same as $n > 3$. The groups in an $A_n{}^3$-cohomology system, H, will often be defined in terms of cochain groups C^r $(r = 0, 1, \cdots, C^r = 0$ if $r > n + 3)$. These are free abelian groups of finite rank, which are related by a coboundary operator $\delta : C^r \to C^{r+1}$, such that $\delta\delta = 0$. By the usual procedure, the r-dimensional cohomology groups $H^r(m)$ with integers reduced mod m can be defined. For integers $p > 0$, $q \geqq 0$ two operators Δ_q, $\mu_{p,q}$ are defined such that $\Delta_q : H^r(q) \to H^{r+1}$, $\mu_{p,q} : H^r(q) \to H^r(p)$. Let $x \in H^r(q)$ and let $x' \in x$. That is to say, x' is a cocycle mod q. Then $\delta x' = qy'$, where y' is an $(r+1)$-absolute cocycle. We define $\Delta_q x = y$, a cohomology class containing y'. Let $c = (p, q)$, then $(p/c)x'$ is a cocycle mod p, and we define $\mu_{p,q}x$ as its cohomology class mod p. It is easily verified that $\Delta_q x$ and $\mu_{p,q}x$ depend only on $x \in H^r(q)$ and not on the particular choice of $x' \in x$. (See J. H. C. Whitehead [3]).

Let H, $H^\dagger$ be $A_n{}^3$-cohomology systems, which are defined in terms of cochain groups $C = \{C^r\}$, $C^\dagger = \{C^{\dagger r}\}$. A cochain map $g : C \to C^\dagger$, obviously induces a (Δ, μ)-homomorphism $f : H \to H^\dagger$. We shall say that f is realized by g. Then

LEMMA 1. (Whitehead [3], [4]) *Any* (Δ, μ)*-homomorphism can be realized by a cochain map.*

3. The main theorem. Let K be an $A_n{}^3$-complex, so that the cohomology groups of K or the cohomology ring of K (in the sense of J. H. C. Whitehead) and homomorphisms defined below form an $A_n{}^3$-cohomology system. Among the homomorphisms, Δ_q and $\mu_{p,q}$ are the same as defined in the last part of section 2, and γ_r is the Steenrod square, $Sq^2 : H^r(2) \to H^{r+2}(2)$ $(r = n, n+1)$. Next we define Φ_i as follows.

Let $x \in N_1 = (Sq^2\mu_{2,4})^{-1}(0)$ and $x' \in x$. Then $Sq^2(\mu_{2,4}x') = \delta v \pmod 2$ and let $x^\dagger = \frac{1}{2}(x' \underset{n}{\cup} x' - x')$ and $\delta x^\dagger = x' \underset{n-1}{\cup} x' \pmod 2$ and $Sq^3Sq^1x = \delta\mathfrak{z}^3x^\dagger \pmod 2$, where $\mathfrak{z}^3x^\dagger = x^\dagger \underset{n-3}{\cup} x^\dagger + x^\dagger \underset{n-2}{\cup} \delta x^\dagger$. $Sq^2Sq^2x' = \delta\mathfrak{z}^2v \pmod 2$, where $\mathfrak{z}^2v = v \underset{n-1}{\cup} v + v \underset{n}{\cup} \delta v$. Adem defined in [9] a cochain map

$$E_i : C^q(K^4) \to C^{q-i}(K)$$

and proved the formula $\delta E_{3n-3}(x'^4) = Sq^2Sq^2x' + Sq^3Sq^1x' \pmod 2$. From these facts we have the operation

$$\Phi_1(x) = E_{3n-3}(x'^4) + \mathfrak{z}^2v + \mathfrak{z}^3x \qquad \bmod Sq^2H^{n+1}(2).$$

And similarly if $x \in (Sq^2)^{-1}(0)$

$$\Phi_2(x) = 2E_{3n-3}(x'^4) + 2\mathfrak{z}^2v + \mathfrak{z}^2\Delta_2x' \qquad \bmod \mu_{4,2}Sq^2H^{n+1}(2).$$

If $Sq^2\Delta_2x \sim 0 \pmod 2$, then it is easily verified that

$$\Phi_2(x) \subset 2H^{n+3}(4)/\mu_{4,2}Sq^2H^{n+1}(2).$$

They are obviously homomorphisms and depend only on $x \in N_1$ (or N_2). (For these homomorphisms, cf. [9], [11].) We now state the main theorem.

THEOREM 1. *Any $A_n{}^3$-cohomology system is properly isomorphic to the cohomology system of some $A_n{}^3$-complex.*

THEOREM 2. *Two $A_n{}^3$-complexes are of the same homotopy type if and only if their cohomology systems are properly isomorphic.*

THEOREM 3. *Let K, L be $A_n{}^3$-complexes, then any proper homomorphism $f^* : H(L) \to H(K)$ has a geometrical realization $f : K \to L$.*

It is easily seen that Theorem 3 implies Theorem 2. (Cf. J. H. C. Whitehead [3].) Hence we shall prove Theorems 1 and 3.

4. $\pi_{n+2}(K^{n+2})$. It is known (S. C. Chang [6]) that any $A_n{}^2$-complex ($n \geqq 3$) has the same homotopy type as a union of elementary polyhedra with a single point in common. Here we shall write down the elementary polyhedra:

$B^r{}_1 = S^r$ $(r = n, n+1, n+2)$.

$B_2(\sigma) = S^n \cup e^{n+1}$, where e^{n+1} is attached to S^n by a map $f: \partial e^{n+1} \to S^n$ of degree σ, a power of a prime.

$B_3(\tau) = S^{n+1} \cup e^{n+2}$, where e^{n+2} is attached to S^{n+1} by a map $f: \partial e^{n+2} \to S^{n+1}$ of degree τ, a power of a prime.

$B_4 = S^n \cup e^{n+2}$, where e^{n+2} is attached to S^n by an essential map of ∂e^{n+2} onto S^n.

$B_5(2^p) = S^n \cup e^{n+2} \cup e^{n+1} = B_4 \cup e^{n+1}$, where e^{n+1} is attached to $S^n \subset B_4$ by a map $f: \partial e^{n+1} \to S^n$ of degree 2^p.

$B_6(2^q) = (S^n \vee S^{n+1}) \cup e^{n+2}$, where e^{n+2} is attached to $S^n \vee S^{n+1}$ by a map $f: \partial e^{n+2} \to S^n \vee S^{n+1}$ of the form $a + b$, where a denotes an essential map of ∂e^{n+2} onto S^n and b maps ∂e^{n+2} onto S^{n+1} with degree 2^q.

$B_7(2^p, 2^q) = (S^n \vee S^{n+1}) \cup e^{n+2} \cup e^{n+1} = B_6(2^q) \cup e^{n+1}$, where e^{n+1} is attached to $S^n \subset B_6(2^q)$ by a map $f: \partial e^{n+1} \to S^n$ of degree 2^p.

To obtain a reduced $A_n{}^3$-complex, we must compute $\pi_{n+2}(K^{n+2})$ of an $A_n{}^2$-complex K^{n+2} (Cf. [10], [11], [13], [14].).

LEMMA 2. *Let Z denote the group of integers. Then*

i) $\pi_{n+2}(S^n) = Z_2$, $\pi_{n+2}(S^{n+1}) = Z_2$, $\pi_{n+2}(S^{n+2}) = Z$.

ii) $\pi_{n+2}(B_2(\sigma)) = 0$ *if* σ *is odd.*

iii) $\pi_{n+2}(B_2(2)) = Z_4$, *its generator* ξ'_2 *is represented by a map covering* e^{n+1} *essentially and* $2\xi'_2$ *is represented by an essential map* $f: S^{n+2} \to S^n \subset B_2(2)$.

iv) $\pi_{n+2}(B_2(2^p)) = Z_2 + Z_2$ *(if* $p > 1$*) its generators* ξ_2 *and* η_2; ξ_2 *is represented by a map covering* e^{n+1} *essentially, and* η_2 *is represented by a map covering* S^n *essentially.*

v) $\pi_{n+2}(B_3(\tau)) = 0$ *if* τ *is odd.*

vi) $\pi_{n+2}(B_3(2^p)) = Z_2$; *its generator* ξ_3 *is represented by a map covering* S^{n+1} *essentially.*

vii) $\pi_{n+2}(B_4) = Z$ *its generator* ι_4 *is represented by a map covering* e^{n+2} *with degree 2.*

viii) $\pi_{n+2}(B_5(2^p)) = Z + Z_2$, *its generators* ι_5, ξ_5, *where* $2\xi_5 = 0$. ι_5 *is represented by a map covering* e^{n+2} *with degree* 2, *and* ξ_5 *is represented by a map covering* e^{n+1} *essentially.*

ix) $\pi_{n+2}(B_6(2^q)) = Z_2$; *its generator* ξ_6 *is represented by a map covering* S^{n+1} *essentially.*

x) $\pi_{n+2}(B_7(2^p, 2^q)) = Z_2 + Z_2$ *its generators* $\xi_{7,1}$, $\xi_{7,2}$ *where* $\xi_{7,1}$ *is the same as is used in* ix) *and* $\xi_{7,2}$ *is the same as is used in* iv).

LEMMA 3. *If* K *is* m*-connected,* $m \geqq 1$, *and* L *is* n*-connected,* $n \geqq 1$, *then*

$$\pi_i(K \vee L) = \pi_i(K) + \pi_i(L) \text{ for } i \leqq m + n$$
$$\pi_{m+n+1}(K \vee L) = \pi_{m+n+1}(K) + \pi_{m+n+1}(L) + \pi_{m+1}(K) \otimes \pi_{n+1}(L)$$

where the isomorphism of the tensor product is defined by means of the Whitehead product. (This was pointed out by the referee.)

One uses a theorem of G. W. Whitehead [15] and calculates $\pi_i(K \times L, K \vee L)$ by the Künneth Theorem and the Hurewicz Theorem to prove this.

Let K^{n+2} be a reduced $A_n{}^2$-complex described as follows (in all the summations there exist finite terms).

$$\begin{aligned} K^{n+2} = &\sum S^n{}_{1,i} + \sum S^{n+1}{}_{1,i} + \sum S^{n+2}{}_{1,i} + \sum B_{2,i}(2) \\ &+ \sum_{q_i>1} B_{2,i}(2^{q_i}) + \sum_{\sigma_i \text{ odd}} B_{2,i}(\sigma_i) + \sum B_{3,i}(2^{q_i}) \\ &+ \sum_{\tau_i \text{ odd}} B_{3,i}(\tau_i) + \sum B_{4,i} + \sum B_{5,i}(2^{p_i}) + \sum B_{6,i}(2^{q_i}) \\ &+ \sum B_{7,i}(2^{p_i}, 2^{q_i}). \end{aligned}$$

From Lemmas 2 and 3, if $n > 3$, $\pi_{n+2}(K^{n+2})$ is a direct sum of the $(n+2)$-nd homotopy groups of the elementary polyhedra of which K^{n+2} consists. If $n = 3$, $\pi_5(K^5)$ is the direct sum of the 5-th homotopy groups of the elementary polyhedra and of the group W generated by the Whitehead product.

For later uses, we shall write the generators of $\pi_{n+2}(K^{n+2})$ as follows

$$\pi_{n+2}(S^n{}_{1,i}) \ni \eta_{1,i},\ \pi_{n+2}(S^{n+1}{}_{1,i}) \ni \xi_{1,i},\ \pi_{n+2}(S^{n+2}{}_{1,i}) \ni \iota_{1,i},$$
$$\pi_{n+2}(B_{2,i}(2)) \ni \xi'_{2,i},\ \pi_{n+2}(B_{2,i}(2^{q_i})) \ni \xi_{2,i},\ \eta_{2,i}$$
$$\pi_{n+2}(B_{3,i}(2^{q_i})) \ni \xi_{3,i},\ \pi_{n+2}(B_{4,i}) \ni \iota_{4,i},$$
$$\pi_{n+2}(B_{5,i}(2^{p_i})) \ni \iota_{5,i},\ \xi_{5,i},\ \pi_{n+2}(B_{6,i}(2^{q_i})) \ni \xi_{6,i},$$
$$\pi_{n+2}(B_{7,i}(2^{p_i}, 2^{q_i})) \ni \xi_{7,1,i},\ \xi_{7,2,i}.$$

5. Reduced $A_n{}^3$-complex. In order to define a reduced $A_n{}^3$-complex, we first introduce the complex K, which has its $(n+2)$-skeleton K^{n+2} as described at the end of Section 4, and we shall transform K to a simpler type of an $A_n{}^3$-complex, which we shall call a reduced $A_n{}^3$-complex. For this transformation, we use the following Lemmas by J. H. C. Whitehead [5], [2].

LEMMA 4. *If two spaces P, Q are of the same homotopy type, and $f: P \to Q$ is a homotopy equivalence, and if a map $\alpha: \partial e^{n+1} \to P$ is given, then the attached spaces $P \cup e^{n+1}$ and $Q \cup e^{n+1}{}_1$ are of the same homotopy type, where e^{n+1} is attached to P by a map $\alpha: \partial e^{n+1} \to P$ and $e^{n+1}{}_1$ is attached to Q by a map $f \cdot \alpha: \partial e^{n+1}{}_1 \to Q$.*

LEMMA 5. *If X, Y are simply connected CR-spaces (a connected compactum which is an* A. N. R.*), then a map $f: X \to Y$ is a homotopy equivalence if each of the induced maps $H^n(X) \to H^n(Y)$ $(n = 2, 3, \cdots)$ is an isomorphism onto.*

We shall fix K^{n+2}. Let $e^{n+3}{}_j$ $(j = 1, 2, \cdots)$ be principal $(n+3)$-cells of K. Then K is an attached space, i. e., $e^{n+3}{}_j$ are attached to K^{n+2} by maps $f_j: \partial e^{n+3}{}_j \to K^{n+2}$. Hence f_j determines an element of $\pi_{n+2}(K^{n+2})$.

We shall define $\beta: C_{n+3}(K) = \pi_{n+3}(K^{n+3}, K^{n+2}) \to \pi_{n+2}(K^{n+2})$ by the above correspondence; then β is a homotopy boundary homomorphism. We shall now write K as $\{K^{n+2}, \beta e^{n+3}{}_j\}$. Henceforth we assume $n > 3$.

$$\begin{aligned}\text{Let } \beta e^{n+3}{}_j = &\sum a^{1,i}{}_j \iota_{1,i} + \sum a^{4,i}{}_j \iota_{4,i} + \sum a^{5,i}{}_j \iota_{5,i} \\ &+ \sum b^{1,i}{}_j \xi_{1,i} + \sum b^{2,i}{}_j \xi_{2,i} + \sum b^{3,i}{}_j \xi_{3,i} \\ &+ \sum b^{5,i}{}_j \xi_{5,i} + \sum b^{6,i}{}_j \xi_{6,i} + \sum b^{7,1,i}{}_j \xi_{7,1,i} \\ &+ \sum b^{7,2,i}{}_j \xi_{7,2,i} + \sum c^{1,i}{}_j \eta_{1,i} + \sum c^{2,i}{}_j \eta_{2,i} + \sum d^{2,i}{}_j \xi'_{2,i},\end{aligned}$$

where the a are integers, the b, c are integers mod 2 and the d are integers mod 4. Then we shall write $\{K^{n+2}, \beta e^{n+3}{}_j\}$ by the matrix form

$$(*) \qquad e^{n+3}{}_j \ [a \ b \ c \ d].$$

Now we consider homotopy equivalences of K^{n+2} onto K^{n+2}. Using Lemma 4, we shall simplify the matrix (*). From Lemma 5 it is easily verified that all the following maps are homotopy equivalences of K^{n+2} onto K^{n+2}.

I. 1) $S^{n+2}{}_{1,i}$ is divided into two hemispheres and the one is mapped onto $S^{n+2}{}_{1,i}$ with degree 1 and the other is mapped onto $S^{n+2}{}_{1,j}$ $(i \neq j)$ with any degree, or onto $S^{n+1}{}_{\lambda,i}$ $(\lambda = 1, 6, 7)$ essentially, or onto $S^{n+1}{}_{3,j}(2^q)$ essentially, or onto $e^{n+1}{}_{2,j}(2^q)$ essentially, or

onto ${e^{n+1}}_{\lambda,j}$ $(\lambda = 5, 7)$ essentially. The remaining part of K^{n+2} is mapped identically.

I. 2) ${S^{n+2}}_{1,i}$ is divided into two hemispheres and the one is mapped onto ${S^{n+2}}_{1,i}$ with degree 1 and the other is mapped onto ${S^n}_{1,j}$ essentially, or onto ${S^n}_{2,j}(2^q)$ essentially. The part of K^{n+2} excluding ${S^{n+2}}_{1,i}$ is mapped identically.

I. 3) An $(n+1)$-cell or an $(n+1)$-sphere belonging to K^{n+2} is divided into two parts and the one is mapped onto itself with degree 1, and the other is mapped onto any n-sphere of K^{n+2} essentially. The cells of K^{n+2} excluding the $(n+1)$-cell or the n-sphere are mapped onto themselves identically.

I. 4) $B_{2,i}(2)$ is divided into two parts and the one is mapped onto $B_{2,i}(2)$ with degree 1 and the other is mapped onto $B_{2,j}(2)$ $(i \neq j)$ with degree 1. The cells of K^{n+2} excluding $B_{2,i}(2)$ are mapped identically. (The reader should be referred to the following operations for detailed accounts.)

Using Lemma 4 and Lemma 2, we shall translate the above homotopy equivalences to allowable operations in the matrix (*) which do not alter the homotopy type of K. That is to say, without change of the homotopy type of K and K^{n+2}, we shall simplify the matrix which represents attaching maps of e^{n+3}.

The following operations are allowable

II. 1) Any two rows or columns are exchangeable.

II. 2) Any column can be added to any other column.

II. 3) The elements of any column can have their signs changed.

II. 4) ${\iota}_{1,i}$'s row can be added to ${\iota}_{1,j}$'s row $(i \neq j)$.

II. 5) ${\iota}_{1,i}$'s row is reduced mod 2 and is added to ξ's row.

II. 6) ${\iota}_{1,i}$'s row is reduced mod 2 and is added to η's row.

II. 7) ${\iota}_{1,i}$'s row is reduced mod 2 and is added to ξ'''s row by considering the non-zero element mod 2 is the generator of the group of integers mod 4.

II. 8) ${\iota}_{1,i}$'s row is reduced mod 2 and is added to ξ'''s row by considering

that the non-zero element mod 2 is the element of order 2 of the group of integers mod 4.

II. 9) ξ's row is added to η's row.

II. 10) ξ's row is added to ξ'''s row considering the non-zero element mod 2 is the element of order 2 of mod 4.

II. 11) ξ'''s row is reduced mod 2 and is added to η's row.

II. 12) $\xi'_{2,i}$'s row is added to $\xi'_{2,j}$'s row $(i \neq j)$.

II. 1)-II. 3) are trivial operations; II. 4) 5) 7) are based on I. 1); II. 6) 8) are based on I. 2); II. 9)-II. 11) are based on I. 3); and II. 12) is based on I. 4).

$$e^{n+3}{}_j \begin{pmatrix} \iota & \xi & \eta & \xi' \\ a & b & c & d \end{pmatrix} \xrightarrow[\text{II, 1, 2, 4)}]{} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & \xi' \\ E_1 & * & * & * & * \\ E_2 & * & * & * & * \end{matrix}$$

$$\xrightarrow[\text{II, 5, 6, 7, 8)}]{} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & \xi' \\ E_1 & * & 0 & 0 & 0 \\ E_2 & * & * & * & * \end{matrix} \xrightarrow[\text{II,1, 2, 3, 12)}]{} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \\ E_1 & * & 0 & 0 & 0 & 0 \\ E_2 & * & * & * & E & 0 \\ & & & & 0 & F \\ & & & & 0 & 0 \end{matrix}$$

$$\xrightarrow[\text{II, 9)}]{} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \\ E_1 & * & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 \\ E_2 & * & * & * & 0 & F \\ & & * & * & 0 & 0 \end{matrix} \xrightarrow[\text{II, 1, 2)}]{} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \\ E_1 & * & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 \\ & & * & * & 0 & F \\ E_2 & * & & & & \\ & & G & * & 0 & 0 \\ & & 0 & * & 0 & 0 \end{matrix}$$

$$\xrightarrow[\text{II, 9)}]{} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \\ E_1 & * & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 \\ & & * & * & 0 & F \\ E_2 & * & & & & \\ & & G & 0 & 0 & 0 \\ & & 0 & * & 0 & 0 \end{matrix} \xrightarrow[\text{II, 1, 2)}]{} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \\ E_1 & * & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 \\ & & G & * & 0 & E_3 \\ E_2 & * & 0 & * & 0 & E_3 \\ & & G & 0 & 0 & 0 \\ & & 0 & * & 0 & 0 \end{matrix}$$

$$\begin{array}{c} \\ \xrightarrow{\text{II, 9, 10)}} \end{array}\begin{array}{c} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \end{matrix} \\ \begin{bmatrix} E_1 & * & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 \\ & & G & 0 & 0 & 0 \\ E_2 & * & 0 & * & 0 & E_3 \\ & & G & 0 & 0 & 0 \\ & & 0 & * & 0 & 0 \end{bmatrix} \end{array} \xrightarrow{\text{II, 1, 2)}} \begin{array}{c} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \end{matrix} \\ \begin{bmatrix} E_1 & * & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 \\ & & G & 0 & 0 & 0 \\ E_2 & * & & & & \\ & & 0 & 0 & 0 & 0 \\ & & 0 & * & 0 & E_3 \\ & & 0 & * & 0 & 0 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ \xrightarrow{\text{II, 1, 2)}} \end{array}\begin{array}{c} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \end{matrix} \\ \begin{bmatrix} E_1 & * & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 \\ & & G & 0 & 0 & 0 \\ E_2 & * & 0 & G & 0 & E_3 \\ & & 0 & 0 & 0 & E_3 \\ & & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \xrightarrow{\text{II, 1, 2, 3)}} \begin{array}{c} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' \end{matrix} \\ \begin{bmatrix} E_1 & * & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 \\ & & G & 0 & 0 & 0 \\ E_2 & * & 0 & G & 0 & E_3 \\ & & 0 & 0 & 0 & F \\ & & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ \xrightarrow{\text{II, 2)}} \end{array}\begin{array}{c} \begin{matrix} \iota_1 & \iota_4\iota_5 & \xi & \eta & & \xi' & \end{matrix} \\ \begin{bmatrix} E_1 & * & 0 & 0 & 0 & 0 & 0 \\ & & * & 0 & E & 0 & 0 \\ E_2 & * & G & 0 & 0 & 0 & 0 \\ & & 0 & G & 0 & 0 & E_3 \\ & & 0 & 0 & 0 & 2E & 0 \\ & & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{l} \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} \begin{matrix} s_0 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{matrix} \end{array} \qquad (\#)$$

where the diagonal elements of E_1 are odd integers and the other elements of E_1 are even; all the elements of E_2 and E_3 are even; the diagonal elements of F are 2 and the other elements of F are zero; the diagonal elements of G are 1 and the part of G lower and left to the diagonal consists of 0.

We can now assume that any $A_n{}^3$-complex has the same homotopy type as the complex K with a matrix of the form (#). We call such a complex a reduced $A_n{}^3$-complex. By the cell of i-th kind, we shall mean an $(n+3)$-cell $e^{n+3}{}_j$ ($\sum_{k<i-1} s_k < j \leqq \sum_{k\leqq i} s_k$). In case $n=3$, the argument about $\pi_5(K^5)$ of section 5 shows that, if we neglect the direct summand group generated by the Whitehead product, the above discussion is still valid.

6. Cohomology operations in a reduced $A_n{}^3$-complex. Let K be a reduced $A_n{}^3$-complex $(n>3)$. Let

$$\beta e^{n+3}{}_j = \sum a^{1,i}{}_{j\iota_{1,i}} + \sum a^{4,i}{}_{j\iota_{4,i}} + \sum a^{5,i}{}_{j\iota_{5,i}}$$

$$(j = 1, \cdots, s_0, s_0+s_1+s_2+s_3+1, \cdots, s_0+s_1+s_2+s_3+s_4)$$

$$\beta e^{n+3}{}_j = \sum a^{1,i}{}_j \iota_{1,i} + \sum a^{4,i}{}_j \iota_{4,i} + \sum a^{5,i}{}_j \iota_{5,i} + \sum b^{1,i}{}_j \xi_{1,i}$$
$$+ \sum b^{2,i}{}_j \xi_{1,i} + \sum b^{3,i}{}_j \xi_{3,i} + \sum b^{5,i}{}_j \xi_{5,i} + \sum b^{6,i}{}_j \xi_{6,i}$$
$$+ \sum b^{7,1,i}{}_j \xi_{7,1,i} + \sum b^{7,2,i}{}_j \xi_{7,2,i} + \sum d^{2,i}{}_j \xi'_{2,i},$$
$$(j = s_0 + 1, \cdots, s_0 + s_1), \text{ where } d^{2,i}{}_j = 0 \text{ or } 1.$$

$$\beta e^{n+3}{}_j = \sum a^{1,i}{}_j \iota_{1,i} + \sum a^{4,i}{}_j \iota_{4,i} + \sum a^{5,i}{}_j \iota_{5,i} + \sum c^{1,i}{}_j \eta_{1,i}$$
$$+ \sum c^{2,i}{}_j \eta_{2,i} + \sum d^{2,i}{}_j \xi'_{2,i}$$
$$(j = s_0 + s_1 + 1, \cdots, s_0 + s_1 + s_2),$$

where $d^{2,i}{}_j$ is even.

$$\beta e^{n+3}{}_j = \sum a^{1,i}{}_j \iota_{1,i} + \sum a^{4,i}{}_j \iota_{4,i} + \sum a^{5,i}{}_j \iota_{5,i} + \sum \mathrm{d}^{2,i}{}_j \xi'_{2,i}$$
$$(j = s_0 + s_1 + s_2 + 1, \cdots, s_0 + s_1 + s_2 + s_3),$$

where $d^{2,i}{}_j$ is even.

Then

i) $\delta S^{n+2}{}_{1,i} = \sum a^{1,i}{}_j e^{n+3}{}_j$, $\delta e^{n+2}{}_{4,i} = \sum 2a^{4,i}{}_j e^{n+3}{}_j$, $\delta e^{n+2}{}_{5,i} = \sum 2a^{5,i}{}_j e^{n+3}{}_j$.

ii) Let j run over all the indices of the cells of the first kind.

$$Sq^2 S^{n+1}{}_{1,i} = \sum b^{1,i}{}_j e^{n+3}{}_j, \qquad Sq^2 e^{n+1}{}_{2,i}(2^{p_i}) = \sum b^{2,i}{}_j e^{n+3}{}_j,$$
$$Sq^2 S^{n+1}{}_{3,i}(2^{q_i}) = \sum b^{3,i}{}_j e^{n+3}{}_j, \qquad Sq^2 e^{n+1}{}_{5,i} = \sum b^{5,i}{}_j e^{n+3}{}_j,$$
$$Sq^2 S^{n+1}{}_{6,i} = \sum b^{6,i}{}_j e^{n+3}{}_j, \qquad Sq^2 S^{n+1}{}_{7,i} = \sum b^{7,1,i}{}_j e^{n+3}{}_j,$$
$$Sq^2 e^{n+1}{}_{7,2,i} = \sum b^{7,2,i}{}_j e^{n+3}{}_j,$$

iii) $$Sq^2 e^{n+1}{}_{2,i}(2) = \sum d^{2,i}{}_j e^{n+3}{}_j.$$

In all the cases above, we think of $S^{n+1}{}_{1,i}, \cdots, e^{n+1}{}_{2,i}(2)$ as cocycles mod 2, and $e^{n+3}{}_j$ as cocycles mod 2.

From the matrix (#) of section 5, it is easily seen that the cocycles mod 2 consisting of the cells of the first kind form a base of $Sq^2 H^{n+1}(K, Z_2)$.

Thus, the cells of the i-th kind $(i = 2, 3, 4)$ form a base of

$$H^{n+3}(K, Z_2)/Sq^2 H^{n+1}(K, Z_2).$$

Hence we can write

iv) $\Phi_1(S^n{}_{1,i}) = \sum c^{1,i}{}_j e^{n+3}{}_j$, $\Phi_1(S^n{}_{2,i}(2^p)) = \sum c^{2,i}{}_j e^{n+3}{}_j$, where j runs over all the indices of the cells of the second kind.

$$\Phi_1(S^n{}_{2,i}(2^{p_i})) = \sum c^{2,i}{}_j e^{n+3}{}_j.$$

Similarly, in $2H^{n+3}(K, Z_4)$, the cells of the second kind form a base of $\mu_{4,2} Sq^2 H^{n+1}(K, Z_2)$. Then, the cells of the i-th kind $(i = 2, 3, 4)$ form a base of $2H^{n+3}(K, Z_4)/\mu_{4,2} Sq^2 H^{n+1}(K, Z_2)$. Hence

v) $\Phi_2(S^n{}_{2,i}(2)) = \sum 1/2 d^{2,i}{}_j e^{n+3}{}_j$, where j runs over all the indices of the cells of the second and the third kind.

To prove ii)-v), the usual method is applicable. That is to say, using an appropriate mapping of K into $M = S^{n+1} \cup e^{n+3}$, where e^{n+3} is attached to S^{n+1} by an essential map of ∂e^{n+3} onto S^{n+1}, or using a map of K into $L = S^n \cup e^{n+3}$, where e^{n+3} is attached to S^n by an essential map of ∂e^{n+3} onto S^n, ii)-v) can be proved.

The above relations are still valid in case $n = 3$. (For detailed accounts, the reader is referred to [9], [10], [11].)

In case $n = 3$, we read the relation between cup products and Whitehead's products. Let us introduce a lexicographic order in the S^n of K^{n+2} by the subscripts of the S^n and let $Z(S^n{}_{k,i}, S^n{}_{k',i'})$ be a cyclic group of order $d = (g, h)$, where g (or h) is the order of $S^n{}_{k,i}$ (or $S^n{}_{k',i'}$) when $S^n{}_*$ is considered as an element of $H_n(K, Z)$. Then the direct summand W generated by Whitehead's products of $\pi_{n+2}(K^{n+2})$ is a direct sum

$$\sum_{(k,i)>(k',i')} Z(S^n{}_{k,i}, S^n{}_{k',i'}).$$

Let ϕ be a projection of $\pi_{n+2}(K^{n+2})$ onto W, and let

$$\phi \circ \beta e^{n+3}{}_j = \sum_{(k,i)>(k',i')} m^{(k,i;k',i')}{}_j [S^n{}_{k,i}, S^n{}_{k',i'}]$$

where $[S^n{}_*, S^n{}_*]$ is a generator of $Z(S^n{}_*, S^n{}_*)$, and $m^{(*,*)}{}_j$ is an element of $Z(S^n{}_*, S^n{}_*)$.

Then

vi) $(S^n{}_{k,i}) \cdot (S^n{}_{k',i'}) = \sum m^{(k,i;k',i')}{}_j e^{n+3}{}_j, \quad (k, i) > (k', i')$.

For the proof of these facts one uses the complex $S^3 \times S^3$.

7. Proof of Theorem 1. $(n > 3)$. Let $H = \{H^n, H^{n+1}, H^{n+2}, H^{n+3}, H^n(2), H^{n+1}(2), H^{n+2}(2), H^{n+3}(2), H^n(4), H^{n+3}(4), \Delta, \mu, \gamma_n, \gamma_{n+1}, \Phi_1, \Phi_2\}$ be an $A_n{}^3$-cohomology system. We shall construct an $A_n{}^3$-complex K, which has an $A_n{}^3$-cohomology system that is properly isomorphic to a given H. We assume (without loss of generality) that $\gamma_n : H^n(2) \to H^{n+2}(2)$ has a normal form, i. e. if γ_n is expressed in terms of a matrix form, each row or column contains at most one unity. (Cf. S. C. Chang and J. H. C. Whitehead [7].)

Let $u^r{}_1, \cdots, u^r{}_{p_r}$ be generators of the summand of H^r whose orders are 0, and let $v^r{}_1, \cdots, v^r{}_{q_r}$ be a generator of the summands of H^r whose orders are powers of 2, and let $W^r{}_1, \cdots, w^{rt}{}_r$ be a generator of the summands of H^r whose orders are powers of odd primes. Let

$$
\begin{array}{c}
 & u^{n+2}{}_i & & & v^{n+2}{}_i & & & v^{n+3}{}_i & & \\
 & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & E & 0 & 0 \\ 0 & 0 & 0 & E & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & E & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & E & 0 \\ 0 & 0 & 0 & 0 & E & 0 & 0 & 0 & 0 \end{pmatrix} & \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{matrix} \\
u^n{}_i & & & & & & & & & \\
v^{n+1}{}_i & & & & & & & & & \\
 & c_1 & & & c_2 & & & c_3 & &
\end{array}
$$

be a matrix representing $\gamma_n\colon H^n(2) \to H^{n+2}(2)$. Let

$$
\begin{aligned}
K^{n+2} = &\sum {}^{a_1}S^n{}_{1,i} + \sum {}^{p_{n+1}}S^{n+1}{}_{1,i} + \sum {}^{c_1+c_3+t_{n+3}}S^{n+2}{}_{1,i} \\
&+ \sum {}^{b_1}B_{2,i}(2^{k_i}) + \sum {}^{t_{n+1}}B_{2,i}(\sigma_i) + \sum {}^{c_2}B_{3,i}(2^{s_i}) \\
&+ \sum {}^{t_{n+2}}B_{3,i}(\tau_i) + \sum {}^{a_2+a_3}B_{4,i} + \sum {}^{b_2+b_3}B_{5,i}(2^{m_i}) \\
&+ \sum {}^{a_4}B_{6,i}(2^{n_i}) + \sum {}^{b_4}B_{7,i}(2^{f_i}, 2^{g_i}).
\end{aligned}
$$

where 2^{k_i} is the order of $v^{n+1}{}_i$, σ_i is the order of $w^{n+1}{}_i$, 2^{s_i} is the order of $v^{n+2}{}_{a_4+b_4+i}$, τ_i is the order of $w^{n+2}{}_i$, 2^{m_i} is the order of $v^{n+1}{}_{b_1+i}$, 2^{n_i} is the order $v^{n+2}{}_i$, 2^{f_i} is the order of $v^{n+1}{}_{b_1+b_2+b_3+i}$, and 2^{g_i} is the order of $v^{n+2}{}_{a_4+i}$. Let

$$
\begin{aligned}
&\gamma_{n+1}S^{n+1}{}_{1,i} = \sum b^{1,i}{}_j e^{n+3}, \qquad \gamma_{n+1}e^{n+1}{}_{2,i}(2^{p_i}) = \sum b^{2,i}{}_j e^{n+3}{}_j, \\
&\gamma_{n+1}S^{n+1}{}_{3,i}(2^{q_i}) = \sum b^{3,i}{}_j e^{n+3}{}_j, \qquad \gamma_{n+1}e^{n+1}{}_{5,i} = \sum b^{5,i}{}_j e^{n+3}{}_j, \\
&\gamma_{n+1}S^{n+1}{}_{6,i} = \sum b^{6,i}{}_j e^{n+3}{}_j, \qquad \gamma_{n+1}S^{n+1}{}_{7,i} = \sum b^{7,1,i}{}_j e^{n+3}{}_j, \\
&\gamma_{n+1}e^{n+1}{}_{7,2,i} = \sum b^{7,2,i}{}_j e^{n+3}{}_j, \qquad \gamma_{n+1}e^{n+1}{}_{2,i}(2) = \sum{}' d^{2,i}{}_j e^{n+3}{}_j,
\end{aligned}
$$

where $e^{n+3}{}_j$ $(j = 1, \cdots, t_{n+3} + q_{n+3} + p_{n+3})$ corresponds with the generator of H^{n+3}. Now we shall modify $\Phi_1\colon N_1 = (\gamma_n\mu_{2,4})^{-1}(0) \to H^{n+3}(2)/\gamma_{n+1}H^{n+1}(2)$ as follows. As $\gamma_{n+1}H^{n+1}(2)$ is a direct summand of $H^{n+3}(2)$. There is a map $\Phi^\dagger{}_1\colon N_1 \to H^{n+3}(2)$ which induces Φ_1. Next $\mu_{4,2}\gamma_{n+1}H^{n+1}(2)$ is a direct summand of $2H^{n+3}(4)$ and $N_2 = \gamma^{-1}{}_n(0) \cap (\gamma_{n+1}\Delta_2)^{-1}(0) \cap N$ is a direct summand of $N^\dagger{}_2 = \gamma^{-1}{}_n(0) \cap (\gamma_{n+1}\Delta_2)^{-1}(0)$, so there is a map $\Phi^\dagger{}_2\colon N^\dagger{}_2 \to 2H^{n+3}(4)$ which induces Φ_2. Let

$$
\Phi^\dagger{}_1 S^n{}_{1,i} = \sum c^{1,i}{}_j e^{n+3}{}_j, \quad \Phi^\dagger{}_1 S^n{}_{2,i}(2^p) = \sum c^{2,i}{}_j e^{n+3}{}_j, \quad \Phi^\dagger{}_2 S^n{}_{2,i}(2) = \sum{}'' d^{2,i}{}_j e^{n+3}{}_j.
$$

Then we shall attach $e^{n+3}{}_j$ $(j = 1, \cdots, t_{n+3} + q_{n+3} + p_{n+3})$ to K^{n+2} as follows.

i) $\beta e^{n+3}{}_j = a^1{}_{jt_1,j}$ $\qquad (j = 1, \cdots, t_{n+3})$

where $a^1{}_j$ is the order of $w^{n+3}{}_j$.

ii) $$\beta e^{n+3}{}_j = a^1{}_j \iota_{1,j} + \sum b^{1,i}{}_j \xi_{1,i} + \sum b^{2,i}{}_j \xi_{2,i} + \sum b^{3,i}{}_j \xi_{3,i} + \sum b^{5,i}{}_j \xi_{5,i} + \sum b^{6,i}{}_j \xi_{6,i} + \sum b^{7,1,i}{}_j \xi_{7,1,i} + \sum b^{7,2,i}{}_j \xi_{7,2,i} + \sum c^{1,i}{}_j \eta_{1,i} + \sum c^{2,i}{}_j \eta_{2,i} + \sum d^{2,i}{}_j \xi'_{2,i}$$

$(j = t_{n+3} + 1, \cdots, t_{n+3} + c_3)$ where $a^1{}_j$ is the order of $v^{n+3}{}_{a_3+b_3+j-t_{n+3}}$, and $d^{2,i}{}_j = {}'d^{2,i}{}_j + 2''d^{2,i}{}_j$.

iii) $$\beta e^{n+3}{}_j = a^4{}_j \iota_{4,j-t_{n+3}-c_3} + \sum b^{1,i}{}_j \xi_{1,i} + \sum b^{2,i}{}_j \xi_{2,i} + \sum b^{3,i}{}_j \xi_{3,i} + \sum b^{5,i}{}_j \xi_{5,i} + \sum b^{6,i}{}_j \xi_{6,i} + \sum b^{7,1,i}{}_j \xi_{7,1,i} + \sum b^{7,2,i}{}_j \xi_{7,2,i} + \sum c^{1,i}{}_j \eta_{1,i} + \sum c^{2,i}{}_j \eta_{2,i} + \sum d^{2,i}{}_j \xi'_{2,i}$$

$(j = t_{n+3} + c_3 + 1, \cdots, t_{n+3} + c_3 + a_3)$ where $a^4{}_j$ is a half of the order of $v^{n+3}{}_{j-t_{n+3}-c}$, and $d^{2,i}{}_j = {}'d^{2,i}{}_j + 2''d^{2,i}{}_j$.

iv) $$\beta e^{n+3}{}_j = a^5{}_j \iota_{5,j-t_{n+3}-c_3-a_3} + \sum b^{1,i}{}_j \xi_{1,i} + \sum b^{2,i}{}_j \xi_{2,i} + \sum b^{3,i}{}_j \xi_{3,i} + \sum b^{5,i}{}_j \xi_{5,i} + \sum b^{6,i}{}_j \xi_{6,i} + \sum b^{7,1,i}{}_j \xi_{7,1,i} + \sum b^{7,2,i}{}_j \xi_{7,2,i} + \sum c^{1,i}{}_j \eta_{1,i} + \sum c^{2,i}{}_j \eta_{2,i} + \sum d^{2,i}{}_j \xi'_{2,i}$$

$(j = t_{n+3} + c_3 + a_3 + 1, \cdots, t_{n+3} + q_{n+3})$, where $a^5{}_j$ is a half of the order of $v^{n+3}{}_{j-t_{n+3}-c_3}$ and $d^{2,i}{}_j = {}'d^{2,i}{}_j + 2''d^{2,i}{}_j$.

v) $$\beta e^{n+3}{}_j = \sum b^{1,i}{}_j \xi_{1,i} + \sum b^{2,i}{}_j \xi_{2,i} + \sum b^{3,i}{}_j \xi_{3,i} + \sum b^{5,i}{}_j \xi_{5,i} + \sum b^{6,i}{}_j \xi_{6,i} + \sum b^{7,1,i}{}_j \xi_{7,1,i} + \sum b^{7,2,i}{}_j \xi_{7,2,i} + \sum c^{1,i}{}_j \eta_{1,i} + \sum c^{2,i}{}_j \eta_{2,i} + \sum d^{2,i}{}_j \xi'_{2,i}$$

$(j = t_{n+3} + q_{n+3} + 1, \cdots, t_{n+3} + q_{n+3} + p_{n+3})$, where $d^{2,i}{}_j = {}'d^{2,i}{}_j + 2''d^{2,i}{}_j$.

We have finished the construction of an $A_n{}^3$-complex, which has a properly isomorphic $A_n{}^3$-cohomology system, to a given H. The verification is very simple, from the construction.

8. Proof of Theorem 3 ($n > 3$). Let K, L be two $A_n{}^3$-complexes. If $f: H(L) \to H(K)$ is a proper homomorphism of the $A_n{}^3$-cohomology system of L into that of K, we shall construct a map $g: K \to L$ which induces f. Without loss of generality, we can assume that K and L are reduced $A_n{}^3$-complexes.

By Lemma 1, f is realized by a cochain map $f^{\#}: C^*(L) \to C^*(K)$. Let $f_{\#}: C(K) \to C(L)$ be a dual chain map of $f^{\#}$. We shall show that $f_{\#}$ can be realized by a map $g: K \to L$. By the fact that $f: H(L) \to H(K)$ is a proper map, it is clear that $f_{\#} | L^{n+2}$ induces a proper homomorphism of an $A_n{}^2$-cohomology system of L^{n+2} into that of K^{n+2}. So, by the theorem due to J. H. C. Whitehead, $f_{\#} | K^{n+2}$ can be realized by a map $g: K^{n+2} \to L^{n+2}$. To proceed to the next step, we state a lemma due to J. H. C. Whitehead.

LEMMA 6. *Let K, L be simply connected complexes and let e^n be a principal cell, where, $n > 2$. Suppose that $f: C_r(K) \to C_r(L)$ be a chain map such that the map $f \mid C_r(K_0)$ $(r = 0, 1, \cdots)$ can be realized by a cellular map $g_0: K_0 \to L$, where $K_0 = K - e^n$. If $g_0 \beta e^n = \beta f e^n$, then g_0 can be extended to a map $g: K \to L$, which realizes the chain map f.*

Let π be a projection of $\pi_{n+2}(K^{n+2})$ onto its torsion group. Then, by the results of section 6,

$$g\beta - \beta f_{\#} = g\beta\pi - \beta\pi f = g\Gamma - \Gamma f$$

where $\Gamma = \beta\pi$.

By (#) of section 5 and the results of section 6 it is clear that

$$\Gamma e^{n+3}{}_j = 0 \text{ if } j = 1, \cdots, s_0, s_0 + s_1 + s_2 + s_3 + 1, \cdots, s_0 + s_1 + s_2 + s_3 + s_4. \tag{8.1}$$

$$\begin{aligned}\Gamma e^{n+3}{}_j = {} & \sum \xi_{1,i}(Sq^2 S^{n+1}{}_{1,i}) e^{n+3}{}_j + \sum \xi_{2,i}(Sq^2 e^{n+1}{}_{2,i}(2^{p_i})) e^{n+3}{}_j \\ & + \sum \xi_{3,i}(Sq^2 S^{n+1}{}_{3,i}(2^{q_i})) e^{n+3}{}_j + \sum \xi_{5,i}(Sq^2 e^{n+1}{}_{5,i}(Sq^2 e^{n+1}{}_{5,i})) e^{n+3}{}_j \\ & + \sum \xi_{6,i}(Sq^2 e^{n+1}{}_{6,i}) e^{n+3}{}_j + \sum \xi_{7,1,i}(Sq^2 S^{n+1}{}_{7,i}) e^{n+3}{}_j \\ & + \sum \xi_{7,2,i}(Sq^2 e^{n+1}{}_{7,i}) e^{n+3}{}_j + \sum \xi'_{2,i}(Sq^2 e^{n+1}{}_{2,i}(2)) e^{n+3}{}_j \\ & (j = s_0 + 1, \cdots, s_0 + s_1).\end{aligned} \tag{8.2}$$

$$\begin{aligned}\Gamma e^{n+3}{}_j = {} & \sum \eta_{1,i}(\Phi_1(S^n{}_{1,i})) e^{n+3}{}_j + \sum \eta_{2,i}(\Phi_1(S^n{}_{2,i}(2^{p_i}))) e^{n+3}{}_j \\ & + \sum 2\xi'_{2,i}(\Phi_2(S^n{}_{2,i}(2))) e^{n+3}{}_j \\ & (j = s_0 + s_1 + 1, \cdots, s_0 + s_1 + s_2).\end{aligned} \tag{8.3}$$

$$\begin{aligned}\Gamma e^{n+3}{}_j = {} & \sum 2\xi'_{2,i}(\Phi_2(S^n{}_{2,i}(2))) e^{n+3}{}_j \\ & (j = s_0 + s_1 + s_2 + 1, \cdots, s_0 + s_1 + s_2 + s_3).\end{aligned} \tag{8.4}$$

Considering that $f^{\#}$ commutes with δ, Sq^2 and Φ_i and that the cochains of the 0-th, the first, the second and the third kind are the images of δ, Sq^2, Φ_1 and Φ_2 respectively, we can see that $f^{\#}$ maps the cochains of the i-th kind of L into cochains of the i-th kind of K $(i = 0, 1, 2, 3)$.

Taking the dual map $f_{\#}$ of $f^{\#}$, we see that $f_{\#}$ maps the chain of the

i-th kind into the sum of chains of the i-th kind and of the fourth kind. So $g\Gamma - \Gamma f_\#$ is zero on any cell of the o-th kind and fourth kind by (8. 1). If $e^{n+3}{}_j$ is a cell of the first kind, it follows from (8. 2) that

$$\begin{aligned} g\Gamma e^{n+3}{}_j &= g(\sum \xi_{1,i}(Sq^2 S^{n+1}{}_{1,i})\, e^{n+3}{}_j + \cdots) \\ &= \sum \xi^\dagger{}_{1,i}(Sq^2 f^\# {}^\dagger S^{n+1}{}_{1,i})\, e^{n+3}{}_j + \cdots \\ &= \sum \xi^\dagger{}_{1,i}(f^\# Sq^2 {}^\dagger S^{n+1}{}_{1,i})\, e^{n+3}{}_j + \cdots \\ &= \sum \xi^\dagger{}_{1,i}(Sq^2 {}^\dagger S^{n+1}{}_{1,i}) f_\# e^{n+3}{}_j + \cdots, \end{aligned}$$

where $\xi^\dagger{}_{1,i}$ and ${}^\dagger S^{n+1}{}_{1,i}$ belong to L. Since $f_\# e^{n+3}{}_j$ is a sum of a cycle of the first kind and a cycle of the fourth kind, we get the following equality by using (8. 1) and (8. 2).

$$\Gamma f_\# e^{n+3}{}_j = \sum \xi^\dagger{}_{1,i}(Sq^2 S^{n+1}{}_{1,i})\, f_\# e^{n+3}{}_j + \cdots.$$

This leads to $g\Gamma - \Gamma f_\# = 0$ on the cells of the first kind. If $e^{n+3}{}_j$ is a cell of the second kind, then

$$\begin{aligned} g\Gamma e^{n+3}{}_j &= g(\sum \eta_{1,i}(\Phi_1(S^n{}_{1,i}))\, e^{n+3}{}_j + \cdots) \\ &= \sum \eta^\dagger{}_{1,i}(\Phi_1(f^\# {}^\dagger S^n{}_{1,i}))\, e^{n+3}{}_j + \cdots \\ &= \sum \eta^\dagger{}_{1,i}(f^\# \Phi_1({}^\dagger S^n{}_{1,i}))\, e^{n+3}{}_j + \cdots \\ &= \sum \eta^\dagger{}_{1,i}(\Phi_1({}^\dagger S^n{}_{1,i}))\, f_\# e^{n+3}{}_j + \cdots. \end{aligned}$$

Since $f_\# e^{n+3}{}_j$ is a sum of a cycle of the second kind and a cycle of the fourth kind, it follows by (8. 1) and (8. 3)

$$\Gamma f_\# e^{n+3}{}_j = \sum{}^\dagger \eta^\dagger{}_{1,i}(\Phi_1({}^\dagger S^n{}_{1,i}))\, f_\# e^{n+3}{}_j + \cdots.$$

Therefore $g\Gamma - \Gamma f_\# = 0$ on the cells of the second kind. This is also true on the cells of the third kind by analogous arguments. This completes the proof.

9. The case $n = 3$. The arguments about $\pi_5(K^5)$ and the results of section 6 enable us to prove Theorem 1 and Theorem 3 in case $n = 3$, as a slight modification of the proof above ($n > 3$).

NAGOYA UNIVERSITY.

BIBLIOGRAPHY.

[1] J. H. C. Whitehead, "On adding relations to homotopy groups," *Annals of Mathematics*, vol. 42 (1941), pp. 409-428.

[2] ———, "On the homotopy type of A. N. R.'s," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 1133-1145.

[3] ———, "On simply connected 4-dimensional polyhedra," *Commentarii Mathematici Helvetici*, vol. 22 (1949), pp. 48-92.

[4] ———, "The homotopy type of a special kind of polyhedron," *Annales de la Societé Polonaise de Mathématique*, vol. 21 (1948), pp. 176-186.

[5] ———, "Combinatorial homotopy," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 213-245.

[6] S. C. Chang, "Homotopy invariants and continuous mappings," *Proceedings of the Royal Society of London*, Series A, vol. 202 (1950), pp. 253-263.

[7] S. C. Chang and J. H. C. Whitehead, "Note on cohomology systems," *Quarterly Journal of Mathematics* (Oxford), (2), vol. 2 (1951), pp. 167-174.

[8] M. E. Steenrod, "Products of cocycles and extensions of mappings," *Annals of Mathematics*, vol. 48 (1947), pp. 290-320.

[9] J. Adem, "The iteration of the Steenrod square in algebraic topology," *Proceedings of the National Academy of Sciences* (U. S. A.), vol. 38 (1952), pp. 720-726.

[10] H. Uehara, "On homotopy type problems of special kinds of polyhedra, I and II," *Osaka Mathematical Journal*, vol. 4 (1952), pp. 145-168 and pp. 169-184.

[11] N. Shimada and H. Uehara, "Some remarks on the Adem Extension Theorem" (to appear shortly).

[12] A. L. Blakers and W. S. Massey, "Homotopy groups of a triad I, II," *Annals of Mathematics*, vol. 53 (1951), pp. 161-205 and vol. 55 (1952), pp. 192-201.

[13] M. G. Barratt and G. F. Paechter, "A note on $\pi_r(V_{n,m})$," *Proceedings of the National Academy of Sciences* (U. S. A.), vol. 38 (1952), pp. 119-121.

[14] P. J. Hilton, "Calculation of the homotopy groups of A_n^2-polyhedra, I, II," *Quarterly Journal of Mathematics* (Oxford), vol. 1 (1950), pp. 299-309 and vol. 2 (1951), pp. 228-240.

[15] G. W. Whitehead, "A generalization of the Hopf invariant," *Annals of Mathematics*, vol. 51 (1950), pp. 192-237.

ON RIESZ SUMMABILITY AND SUMMABILITY BY DIRICHLET'S SERIES: A FURTHER ADDENDUM AND CORRIGENDUM.*

By C. T. RAJAGOPAL.

1. Terminology. It is perhaps necessary to explicitly call attention to the terminology of my earlier paper with the same title [4], which is not always in accord with practice elsewhere. Let $A(u)$ be a function of bounded variation in every finite interval of $u \geqq 0$, assumed (for convenience) to be such that $A(0) = 0$. We understand by the Riesz (integral) mean of $A(x)$, of order $r \geqq 0$, the function $\sigma_r(x) = A_r(x)/x^r$, where

$$A_r(x) = r\int_0^x (x-u)^{r-1}A(u)\,du,\ r > 0, \qquad \sigma_0(x) = A_0(x) = A(x).$$

If $A(x)$ is a λ-step function defined with respect to a sequence $0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots$, $\lambda_n \to \infty$, i. e., if $A(x)$ is defined by

$$A(x) = a_1 + a_2 + \cdots + a_n \text{ for } \lambda_n \leqq x < \lambda_{n+1},\ n = 1, 2, 3, \cdots,$$

$A(x) = 0$ for $\lambda_0 \leqq x < \lambda_1$, then $\sigma_r(x)$ is the Riesz typical mean of $\sum a_n$, of order r and type λ_n (or λ). In the special case $\lambda_n = n$, the Riesz typical mean is written $R_r(x)$ and called the Riesz arithmetic mean; it can, for some purposes (e. g. Theorem t' of the next section), take the place of the Cesàro mean of order $r \geqq 0$ of $\sum a_n$ defined by $c_r(n) = [\sum_{\nu=0}^{n} C_{\nu+r,r}a_{n-\nu}]/C_{n+r,n}$. The integral $F(t) = \int_0^\infty e^{-ut}d\{A(u)\}$, $t > 0$, whenever it exists, is called the Laplace transform of $A(u)$. Its special case, when $A(u)$ is the λ-step function already defined, is the Dirichlet series $\sum a_n e^{-\lambda_n t}$ called the $D(\lambda_n)$-transform of $\sum a_n$; the further special case in which $\lambda_n = n$ is the power series $\sum a_n e^{-nt}$ called the Abel transform (or mean) of $\sum a_n$.

$A(u)$ is said to be summable to sum s by the r-th Riesz mean $(r > 0)$ or by the Laplace transform (whether in the general case or in one of the special cases), according as $\sigma_r(u) \to s$ as $u \to \infty$ or $F(t) \to s$ as $t \to +0$. Similarly $\sum a_n$ is said to be summable by the r-th Cesàro mean $(r > 0)$ when $c_r(n) \to s$ as $n \to \infty$.

* Received August 15, 1952, substantially revised March 31, 1953.

The opening statement of my previous paper ([4], p. 371) should be understood to mean: When $A(x)$ is a sequence or series $\sum a_n$, its summability by the k-th Riesz arithmetic mean ($k > 0$) necessarily implies its summability by the Laplace transform in the form of the Abel mean.

2. A theorem on summability by Riesz means. The following theorem appears in the concluding section of my earlier paper ([4], p. 377).

THEOREM t. *Write*

$$\limsup_{x\to\infty} \sigma_r(x) = \bar{\sigma}_r, \quad \liminf_{x\to\infty} \sigma_r(x) = \underline{\sigma}_r,$$

$$\lim_{r\to\infty} \bar{\sigma}_r = \bar{\sigma}_\infty, \quad \lim_{r\to\infty} \underline{\sigma}_r = \underline{\sigma}_\infty.$$

Then $\bar{\sigma}_\infty = \underline{\sigma}_\infty = s$ (*finite*) *ensures* $\bar{\sigma}_r = \underline{\sigma}_r = s$ *for all sufficiently large* r, *say* $r \geqq r_0 + 1$.

The previously known analogue of Theorem t for Cesàro means ([3], p. 199), due to Littlewood, is as follows.

THEOREM t'. *In the case of the series* $\sum a_n$, *let*

$$\limsup_{n\to\infty} c_r(n) = \bar{c}_r, \quad \liminf_{n\to\infty} c_r(n) = \underline{c}_r,$$

$$\lim_{r\to\infty} \bar{c}_r = \bar{c}_\infty, \quad \lim_{r\to\infty} \underline{c}_r = \underline{c}_\infty.$$

Then $\bar{c}_\infty = \underline{c}_\infty = s$ (*finite*) *ensures* $\bar{c}_r = \underline{c}_r = s$ *for all sufficiently large* r, *say* $r \geqq r_0 + 1$.

Pennington ([3], p. 200) has recently pointed out that the proof of Theorem t outlined by me is incorrect,[1] and given an interesting alternative proof along entirely different lines ([3], pp. 201-206). However, it may be pointed out that my proof, with an obvious modification, is valid for Theorem t'. For, the hypothesis of Theorem t' implies (i) that $c_r(n)$ is bounded as $n \to \infty$ for all large r, say $r \geqq r_0$, and (ii) that $\sum a_n e^{-nt}$ converges for $t > 0$ and tends to s as $t \to +0$. The implication (i) has the consequence that $R_r(x)$, as also $B_r(x)/x^{r+1} \equiv (r+1)\{R_r(x) - R_{r+1}(x)\}$, is bounded as $x \to \infty$ for $r \geqq r_0$ ([1], p. 113, proof of Theorem 58) and, taken along with (ii),

[1] The mistake is in the assumption that the boundedness of the Riesz (integal) mean $\sigma_r(x)$ as $x \to \infty$, for some $r > 0$, ensures the existence of the Laplace transform $F(t)$ for $t > 0$. Not even the convergence of $\sigma_r(x)$ as $x \to \infty$ suffices to ensure the existence of $F(t)$ for $t > 0$. The opening sentence of [4] which has this wrong implication should be corrected as in the last sentence of the first section of the present paper.

establishes the summability of $\sum a_n$ by $R_r(x)$ for $r \geqq r_0 + 1$, when we appeal to Theorem 1 of my previous paper [4]. But summability of $\sum a_n$ by $R_r(x)$ implies summability by $c_r(n)$ ([1], Theorem 58) and so the proof of Theorem t' is complete.

3. Some lemmas and their proofs. The proof of Theorem t' given above can be adapted to prove Theorem t, by means of the following lemmas of which Lemmas III and IV are for alternative and not simultaneous use.

LEMMA I. *If there is an* $r = r_0 \geqq 0$ *such that* $|\sigma_{r_0}(x)| \leqq K$, *then*

(i) $\Phi_{r_0}(t) = t^{r_0+1}/\Gamma(r_0+1)\int_0^\infty e^{-ut}A_{r_0}(u)\,du$ *converges absolutely for* $t > 0$,

(ii) *for any* $k > 1$,

$$K + \liminf_{t\to+0} \Phi_{r_0}(t) \geqq \nu_k^{(r_0)} e^{-(r_0+1)}[K + \underline{\sigma}_{r_0+k}],$$

$$K - \limsup_{t\to+0} \Phi_{r_0}(t) \geqq \nu_k^{(r_0)} e^{-(r_0+1)}[K - \overline{\sigma}_{r_0+k}],$$

where $\nu_k^{(r_0)} = \dfrac{\Gamma(k)}{\Gamma(r_0+k+1)} \dfrac{(r_0+k)^{r_0+k}}{(k-1)^{k-1}}$.

LEMMA II. *If* $\overline{\sigma}_\infty = \underline{\sigma}_\infty = s$ (*finite*), *then we can find* $r_0 > 0$ *so that the conclusions* (i) *and* (ii) *of Lemma* I *hold and furthermore* $\Phi_{r_0}(+0) = s$.

LEMMA III. *If there is an* $r_0 \geqq 0$ *such that* (*in the notation of Lemma* I)

(1) $\Phi_{r_0}(t)$ *converges absolutely for* $t > 0$ *and tends to* s *as* $t \to +0$,

(2) $B_{r_0}(x) \equiv (r_0+1)\{xA_{r_0}(x) - A_{r_0+1}(x)\} = O_L(x^{r_0+1}), \quad x \to \infty,$
then $\lim_{x\to\infty} \sigma_{r_0+1}(x) = s$.

LEMMA IV. *Suppose that* $A(u) \geqq 0$, $A(u)e^{-ut}$ *is integrable over* $(0,\infty)$ *for* $t > 0$ *and* (*in the notation of Lemma* I)

$$t^{-1}\Phi_0(t) = \int_0^\infty e^{-ut}A(u)\,du \sim Ht^{-\alpha}, \qquad t \to +0, \tag{3}$$

where $\alpha \geqq 0$, $H \geqq 0$; *the supposition* (3) *being interpreted to mean* $t^{-1}\Phi_0(t) = o(t^{-\alpha})$ *in the case* $H = 0$. *Then*

$$\int_0^x A(u)\,du \sim Hx^\alpha/\Gamma(\alpha+1), \qquad x \to \infty.$$

Proof of Lemma I. The conclusion (i) of the lemma is obvious.

To prove the conclusion (ii), we shall first suppose that $\sigma_{r_0}(x) \geqq 0$, or $A_{r_0}(x) \geqq 0$. Then, for a given $y > 0$,

$$(4) \quad \frac{y^{r_0+1}}{\Gamma(r_0+1)} \int_0^x e^{-uy} A_{r_0}(u)du = \frac{1}{\Gamma(r_0+1)x^{r_0+k}} \int_0^x h(u)(x-u)^{k-1} A_{r_0}(u)du$$

where

$$(5) \quad h(u) = e^{-uy}x^{r_0+k}y^{r_0+1}/(x-u)^{k-1}.$$

Now $h(u)$ defined by (5) has absolute minimum, corresponding to $u = x - (k-1)/y = U$ (say), whose value is

$$(6) \quad h(U) = e^{-xy}(xy)^{r_0+k}e^{k-1}/(k-1)^{k-1}.$$

Consequently we have, from (4),

$$(7) \quad \frac{y^{r_0+1}}{\Gamma(r_0+1)} \int_0^x e^{-uy} A_{r_0}(u)du \geqq \frac{h(U)}{\Gamma(r_0+1)x^{r_0+k}} \int_0^x (x-u)^{k-1} A_{r_0}(u)du.$$

$h(U)$ in (6), considered as a function of xy, has absolute maximum when $xy = r_0 + k$, this maximum being

$$(8) \quad e^{-(r_0+1)}(r_0+k)^{r_0+k}/(k-1)^{k-1} = \nu_k^{(r_0)}e^{-(r_0+1)}\Gamma(r_0+k+1)/\Gamma(k).$$

In (7) we can choose $x = (r_0+k)/y$ and replace $h(U)$ by its largest value given by (8), getting the best form of (7):

$$(9) \quad \frac{y^{r_0+1}}{\Gamma(r_0+1)} \int_0^x e^{-uy} A_{r_0}(u)du \geqq \nu_k^{(r_0)}e^{-(r_0+1)} \frac{\Gamma(r_0+k+1)}{\Gamma(k)\Gamma(r_0+1)x^{r_0+k}} \int_0^x (x-u)^{k-1} A_{r_0}(u)du.$$

Now, by a well-known formula,

$$(10) \quad A_{r_0+k}(x) = \Gamma(r_0+k+1)/\{\Gamma(k)\Gamma(r_0+1)\} \int_0^x (x-u)^{k-1} A_{r_0}(u)du.$$

Substituting for the integral on the right side of (9) its value in (10), we obtain from (9):

$$(11) \quad \Phi_{r_0}(y) \geqq y^{r_0+1}/\Gamma(r_0+1) \int_0^x e^{-uy} A_{r_0}(u)du \geqq \nu_k^{(r_0)}e^{-(r_0+1)}\sigma_{r_0+k}(x), \qquad x = (r_0+k)/y.$$

Letting $y \to +0$ in (11), we have

$$(12) \quad \liminf_{y\to+0} \Phi_{r_0}(y) \geqq \nu_k^{(r_0)}e^{-(r_0+1)} \liminf_{y\to+0} \sigma_{r_0+k}((r_0+k)/y) = \nu_k^{(r_0)}e^{-(r_0+1)}\sigma_{r_0+k}.$$

If $\sigma_{r_0}(x)$ does not satisfy the condition $\sigma_{r_0}(x) \geqq 0$ but is such that $|\sigma_{r_0}(x)| \leqq K$, or $\sigma_{r_0}(x) \geqq -K$ and $-\sigma_{r_0}(x) \geqq -K$, we can replace $A_{r_0}(u)$ in the preceding argument successively by $Ku^{r_0} + A_{r_0}(u)$ and $Ku^{r_0} - A_{r_0}(u)$ and deduce from (12) the two inequalities of the conclusion (ii) of Lemma II.

Proof of Lemma II. The hypothesis of Lemma II implies that $\sigma_r(x)$ is bounded as $x \to \infty$ for all large r. Hence we can find $r_0 > 0$ such that $|\sigma_{r_0}(x)| \leqq K$ and consequently the conclusions (i) and (ii) of Lemma I hold. The further conclusion $\Phi_{r_0}(t) \to s$ as $t \to +0$ follows the conclusion (ii) of Lemma I when we let $k \to \infty$, since $\nu_k^{(r_0)} \to 1$ as $k \to \infty$.

Proof of Lemma III. This lemma is a slightly modified form of an extension of a theorem of Szász already obtained by me ([4], Theorem 1).[2] It requires no proof if it is recalled that, in my extension just referred to, I have in effect assumed that

$$\int_0^\infty e^{-ut} d\{A(u)\} \text{ converges for } t > 0 \text{ and tends to } s \text{ as } t \to +0,$$

derived from this assumption the hypothesis (1) of Lemma III, this derivation involving an appeal to known theorems ([7], Chapter II, Theorem 2. 3*a* and §§ 8. 1, 8. 2), and finally used (1) and (2) to establish the conclusion of Lemma III ([4], Lemma 4 and § 1. 2).

Proof of Lemma IV. This is omitted, the lemma being a well-known theorem of Hardy and Littlewood ([2], Theorem 1).

4. Proof of Theorem *t*. The hypothesis of Theorem *t*, which is also that of Lemma II, implies (as a result of the lemma) the existence of an $r_0 > 0$ such that $\sigma_{r_0}(x) = O(1)$ as $x \to \infty$ and hence (1) holds. Now $\sigma_{r_0}(x) = O(1)$ also implies (2) in the form $B_{r_0}(x) = O(x^{r_0+1})$. Hence the conclusion of Theorem *t*, which is included in that of Lemma III, follows on an application of Lemma III.

The referee has suggested that Lemma IV may be used instead of Lemma III in the preceding proof, and the proof concluded thus. The condition $\sigma_{r_0}(x) = O(1)$ shows that we can find a constant K such that $A_{r_0}(x) + Kx^{r_0} \geqq 0$; while the condition $\Phi_{r_0}(t) \to s$ as $t \to +0$ implies

$$\int_0^\infty e^{-ut}\{A_{r_0}(u) + Ku^{r_0}\} du \sim (s + K)\Gamma(r_0 + 1)t^{-(r_0+1)}, \qquad t \to +0.$$

[2] In fact Lemma III reduces to Szász's theorem ([5], Theorem 1) when $r_0 = 0$.

Hence we can take, in Lemma IV, $A_{r_0}(u) + Ku^{r_0}$ instead of $A(u)$, $r_0 + 1$ instead of α, and conclude that

$$\int_0^x \{A_{r_0}(u) + Ku^{r_0}\}du \sim (s + K)\Gamma(r_0 + 1)x^{r_0+1}/\Gamma(r_0 + 2), \quad x \to \infty,$$

or

$$A_{r_0+1}(x) = (r_0 + 1)\int_0^x A_{r_0}(u)\,du \sim sx^{r_0+1}, \qquad x \to \infty,$$

whence the conclusion of Theorem t follows at once.

Remarks on the proof of Theorem t. (i) It is clear from the proof of Lemma III ([4], § 1. 2) that the lemma is effectively a transformation of the well-known case $\alpha = 1$ of Lemma IV. Thus it would seem that there is no essential gain in simplicity resulting from the use of Lemma IV in place of Lemma III.

(ii) Pennington remarks, apropos of the version of Theorem t for Riesz typical means, that its proof by Dirichlet's series requires the use of quite deep-lying results in the uniform summability of such series. The proof of Theorem t given here, along the same lines, makes use of the standard techniques employed in my previous paper [4] including Doetsch's 'positive' Tauberian theorem either in its simplest form (case $\alpha = 1$ of Lemma IV) or in a generalized form (Lemma IV).

5. Minor corrigenda and addenda. In [4], p. 374, relation (10), u^{k+1} of the right-hand member is to be corrected to u^{k+2}.

In [4], p. 374, condition (ii) of Corollary 1, $B_k(x) \geqq -k\lambda_n{}^{k+1}$ is to be corrected to $B_k(x) \geqq -K\lambda_n{}^{k+1}$.

In [4], line 6 of p. 851, 'vol. 59' is to be corrected to 'vol. 69.'

In [4], p. 851, condition (i*) of Corollary 3, $t \to \infty$ is to be corrected to $t \to +0$.

The inversion of integrations, in the proofs of Lemmas 4, 5 of [4], may be justified by Fubini's theorem, if we suppose that the Laplace transform $F(t)$ of $A(u)$ is defined as a Stieltjes integral, as in the first section of this paper, so that

$$F(t) \equiv \int_0^\infty e^{-ut}d\{A(u)\} = t^{k+1}/\Gamma(k + 1)\int_0^\infty A_k(u)e^{-ut}du \equiv \Phi_k(t),$$
$$t > 0, k \geqq 0,$$

where the second integral converges absolutely ([7], Chapter II, Theorem 2. 3*a* and §§ 8, 1, 8. 2).

In the proof of Lemma 5 of [4], p. 376, the passage from the step in line 4 to the next step may be elucidatel thus. Since the inner integral in the former step is non-negative, we have

$$\nu_k^{(r)}\Gamma(r+1)\sigma_{r+k}(x) \leqq e^{r+1}y^{r+1}\int_0^\infty e^{-uy}du\int_0^u r(u-t)^{r-1}A(t)\,dt,$$
$$y=(r+k)/x,$$

and this leads to the latter step, by Fubini's theorem.

Lemma 4 of [4] has been recently proved by Szász ([6], § 3) by a somewhat different method.

In conclusion, I would like to thank the referee for helping me to revise thoroughly this paper.

RAMANUJAN INSTITUTE OF MATHEMATICS (KARAIKUDI), MADRAS.

REFERENCES.

[1] G. H. Hardy, *Divergent Series* (Oxford, 1949).

[2] G. H. Hardy and J. E. Littlewood, "Notes on the theory of series (XI): On Tauberian theorems," *Proceedings of the London Mathematical Society*, vol. 30 (1930), pp. 23-37.

[3] W. B. Pennington, "A Tauberian theorem on the oscillation of Riesz means," *Journal of the London Mathematical Society*, vol. 27 (1952), pp. 199-206.

[4] C. T. Rajagopal, "On Riesz summability and summability by Dirichlet's series," *American Journal of Mathematics*, vol. 69 (1947), pp. 371-378, 851-852.

[5] O. Szász, "Converse theorems of summability for Dirichlet's series," *Transactions of the American Mathematical Society*, vol. 39 (1936), pp. 117-130.

[6] ———, "On products of summability methods," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 257-263

[7] D. V. Widder, *The Laplace Transform* (Princeton Mathematical Series, vol. 16, 1946).

AN INTERPOLATION PROBLEM FOR HARMONIC FUNCTIONS.*

By J. L. WALSH.

The study of functions $f(z)$ analytic interior to the unit circle and defined there by requirements of interpolation is due especially to Kakeya, Takenaka, Malmquist, and Walsh [1] if the functions are subjected to the requirement

$$\int_C |f(z)|^2 \, |dz| < \infty, \qquad C: |z| = 1, \tag{1}$$

and is due especially to Lokki [2] if (1) is replaced by the Dirichlet integral of $f(z)$ over $|z| < 1$. It is the object of the present note to indicate that the extremal methods developed by these authors, and the theory of reproducing kernels developed by Bergman [3] and others [4], apply also to the study of real functions $u(z)$ harmonic interior to C and defined there by requirements of interpolation, where $u(z)$ satisfies a condition analogous to (1). We indicate the differences of method required for harmonic rather than analytic functions, and present the specific formulas involved.

We denote by H the class of functions $(z = re^{i\theta})$,

$$u(z) \equiv \tfrac{1}{2}a_0 + \sum_{k=1}^{\infty} r^k(a_k \cos k\theta + b_k \sin k\theta), \qquad \sum_{k=1}^{\infty} (a_k^2 + b_k^2) < \infty, \tag{2}$$

each of which is clearly harmonic in $|z| < 1$. For a given function $u(z)$, the coefficients a_n and b_n are uniquely determined, because on each diameter of $C: |z| = 1$ the function is represented by a power series in the real variable r; such a representation is unique. The series in (2) converges in the mean on C to some function $u(e^{i\theta})$, and since the partial sums are represented in $|z| < 1$ by their Poisson's integral taken over C, it follows that $u(z)$ is also represented in $|z| < 1$ by the Poisson integral of $u(e^{i\theta})$ taken over C. The *norm* $\| u \|$ of $u(z)$ is defined as the (non-negative) square root of

$$\frac{1}{\pi} \int_C [u(e^{i\theta})]^2 d\theta = \tfrac{1}{2}a_0^2 + \sum_{k=1}^{\infty} (a_k^2 + b_k^2). \tag{3}$$

A function of zero norm vanishes identically. The first series in (2) con-

*Received May 12, 1953.

verges uniformly on every circle $|z| = r < 1$, since the a_k and b_k are bounded, and we have

$$\frac{1}{\pi}\int_{|z|=r}[u(re^{i\theta})]^2 d\theta = \tfrac{1}{2}a_0^2 + \sum_{k=1}^{\infty}(a_k^2 + b_k^2)r^{2k}; \tag{4}$$

this last member is not greater than the second member of (3), and by Abel's theorem approaches the second member of (3) as $r \to 1$. Of course (4) is valid for any function $u(z)$ harmonic in $|z| < 1$, so the last condition in (2) can also be interpreted in the form

$$\lim_{r\to 1}\int_0^{2\pi}|u(re^{i\theta})|^2 d\theta < \infty; \tag{5}$$

the integral whose limit appears here increases monotonically with r, as is obvious from (4).

1. An extremal function.

THEOREM 1. *With $|\alpha| < 1$, let $f_\alpha(z)$ be the function of class H which takes on the value unity for $z = \alpha$ whose norm is least. Then for an arbitrary function $u(z)$ of class H we have*

$$u(\alpha) = \int_C u(z)f_\alpha(z)|\,dz\,| \Big/ \int_C f_\alpha(z)|\,dz\,|. \tag{6}$$

From (2) we have ($|a_n\cos n\theta + b_n \sin n\theta|^2 \leqq a_n^2 + b_n^2$)

$$\begin{aligned}|u(re^{i\theta})| &\leqq (\tfrac{1}{2} + \sum_1^{\infty} r^{2n})^{\frac{1}{2}}(\tfrac{1}{2}a_0^2 + \sum_1^{\infty}(a_n^2 + b_n^2))^{\frac{1}{2}} \\ &= \|u\|[\tfrac{1}{2}(1+r^2)/(1-r^2)]^{\frac{1}{2}},\end{aligned} \tag{7}$$

from which it follows that a sequence of functions $u_n(z)$ of class H of bounded norm is normal in $|z| < 1$. A suitable subsequence converges uniformly on any closed set interior to C to some harmonic function $u(z)$. The monotonic character of the integral in (4) implies

$$\int_0^{2\pi}|u(re^{i\theta})|^2\,d\theta = \lim_{n\to\infty}\int_0^{2\pi}|u_n(re^{i\theta})|^2\,d\theta \leqq \pi\|u_n\|^2;$$

since $\|u_n\|$ is bounded, (5) is satisfied and $u(z)$ is of class H, with $\|u\| \leqq \limsup \|u_n\|$. The set of functions of class H taking the value unity in $z = \alpha$ contains the function $u(z) \equiv 1$, hence is not empty; there exists a function $f_\alpha(z)$ of this set of minimum norm. We have $\|f_\alpha\| > 0$ by virtue of $f_\alpha(\alpha) = 1$.

LEMMA 1. *A function $u_0(z)$ of class H is orthogonal to $f_\alpha(z)$ when and only when $u_0(\alpha) = 0$.*

If we have $u_0(\alpha) = 0$, we have for every real ϵ by the definition of $f_\alpha(z)$

$$\| f_\alpha \|^2 \leqq \| f_\alpha + \epsilon u_0 \|^2 = \| f_\alpha \|^2 + 2\epsilon(f_\alpha, u_0) + \epsilon^2 \| u_0 \|^2, \tag{8}$$

where we use the generic notation

$$(f, g) = \frac{1}{\pi} \int_C f(z) g(z) | dz |.$$

Since (8) holds for every ϵ we must have $(f_\alpha, u_0) = 0$. Conversely, if we have $(f_\alpha, u_0) = 0$ but $u_0(\alpha) \neq 0$, the function

$$(f_\alpha(z) + \epsilon u_0(z))/(1 + \epsilon u_0(\alpha))$$

is of class H, takes the value unity when $z = \alpha$, and has as norm the square root of

$$(\| f_\alpha \|^2 + \epsilon^2 \| u_0 \|^2)/[1 + \epsilon u_0(\alpha)]^2,$$

a norm which for a suitably chosen numerically small ϵ is less than that of $f_\alpha(z)$. This contradiction of the definition of $f_\alpha(z)$ establishes Lemma 1.

If $u(z)$ is an arbitrary function of class H, the function $u(z) - u(\alpha)$ is of class H and vanishes for $z = \alpha$, so by Lemma 1 we have

$$\int_C [u(z) - u(\alpha)] f_\alpha(z) | dz | = 0;$$

in particular from $u(z) \equiv f_\alpha(z)$ we have $\pi \| f_\alpha \|^2 = \int_C f_\alpha(z) | dz | \neq 0$, (6) follows, and Theorem 1 is established.

The function $f_\alpha(z)$ of Theorem 1 is unique, for if a second such function $F_\alpha(z)$ exists, the function $f_\alpha(z) - F_\alpha(z)$ vanishes for $z = \alpha$, hence by Lemma 1 is orthogonal to both $f_\alpha(z)$ and $F_\alpha(z)$, hence orthogonal to $f_\alpha(z) - F_\alpha(z)$, and vanishes identically. *The function $f_\alpha(z)$ of class H is uniquely characterized by $f_\alpha(\alpha) = 1$ and the property of being orthogonal to every $u_0(z)$ of class H which vanishes in $z = \alpha$*; for if a second function $F_\alpha(z)$ has these properties the function $f_\alpha(z) - F_\alpha(z)$ is orthogonal to both $f_\alpha(z)$ and $F_\alpha(z)$, hence is orthogonal to $f_\alpha(z) - F_\alpha(z)$ and vanishes identically. If $\Phi(z)$ is a function of class H and if for every $u_0(z)$ in H we have $u_0(\alpha) = \pi(u_0(z), \Phi(z))$, then the function $\Phi(z) - f_\alpha(z)/\pi(f_\alpha(z), 1)$ is orthogonal to every $u_0(z)$ in H, is orthogonal to itself, and vanishes identically; thus we have $\Phi(z) \equiv f_\alpha(z)/\pi(f_\alpha(z), 1)$.

Lagrange's multipliers give a necessary condition for a minimum and thereby enable us to compute $f_\alpha(z)$. In the notation (2) for $f_\alpha(z)$ with $\alpha = \rho e^{i\phi}$ we have to minimize

$$\tfrac{1}{2} a_0^2 + \sum_1^\infty (a_n^2 + b_n^2) \tag{9}$$

subject to the auxiliary condition

$$\frac{1}{2}a_0 + \sum_1^\infty \rho^n (a_n \cos n\phi + b_n \sin n\phi) = 1. \tag{10}$$

We thus write $(\lambda\mu \neq 0)$

$$F(\lambda, \mu) \equiv \lambda[\tfrac{1}{2}a_0 + \sum_1^\infty \rho^n (a_n \cos n\phi + b_n \sin n\phi)] + \mu[\tfrac{1}{2}a_0^2 + \sum_1^\infty (a_n^2 + b_n^2)],$$

$$\partial F/\partial a_n = \lambda\rho^n \cos n\phi + 2\mu a_n = 0, \qquad n > 0;$$
$$\partial F/\partial a_0 = \tfrac{1}{2}[\lambda + 2\mu a_0] = 0,$$
$$\partial F/\partial b_n = \lambda\rho^n \sin n\phi + 2\mu b_n = 0, \qquad n > 0.$$

Substitution of these unique values of a_n and b_n in (10) now yields

$$-(\lambda/2\mu)[\tfrac{1}{2} + \sum_1^\infty \rho^{2n}(\cos^2 n\phi + \sin^2 n\phi)] = 1,$$
$$-(\lambda/2\mu) = 2(1-\rho^2)/(1+\rho^2),$$
$$\tfrac{1}{2}a_0 = (1-\rho^2)/(1+\rho^2), \qquad a_n = 2(1-\rho^2)\rho^n \cos n\phi/(1+\rho^2),$$
$$b_n = 2(1-\rho^2)\rho^n \sin n\phi/(1+\rho^2),$$

$$f_\alpha(re^{i\theta}) \equiv (1-\rho^2)[(1-r^2\rho^2)/(1-2r\rho\cos(\phi-\theta) + r^2\rho^2)]/(1+\rho^2). \tag{11}$$

The denominator in (6) reduces to $2\pi(1-\rho^2)/(1+\rho^2)$, so the second member of (6) reduces to the familiar form of Poisson's integral.

Our entire proof of Theorem 1 can be based directly on Poisson's integral rather than on the general theory of orthogonal functions; the contrast in method is that of derivation rather than immediate verification, methods of general rather than restricted applicability. The computation of $f_\alpha(z)$ must of necessity yield the kernel of Poisson's integral, for we have already shown that any function $\Phi(z)$ of class H with the reproducing property $u_0(\alpha) = \int_C u_0(z)\Phi(z)|\,dz\,|$ for every $u_0(z)$ in H must be identical with $f_\alpha(z)/\int_C f_\alpha(z)|\,dz\,|$. Incidentally, we have here a new proof of the validity of Poisson's formula.

Another form of (11) is $(z = re^{i\theta})$

$$f_\alpha(re^{i\theta}) \equiv (1-\rho^2)\,\mathrm{Re}[(1+\bar{\alpha}z)/(1-\bar{\alpha}z)]/(1+\rho^2), \tag{12}$$

which will prove especially useful.

2. Interpolation and orthogonal functions. We are now in a position to study the general problem of interpolation.

THEOREM 2. *Let the distinct points $\alpha_0, \alpha_1, \alpha_2, \cdots$ interior to C be given,*

finite or infinite in number. Let $\psi_n(z)$ be the function of class H of minimum norm which satisfies the auxiliary conditions

$$(13) \qquad \psi_n(\alpha_0) = \psi_n(\alpha_1) = \cdots = \psi_n(\alpha_{n-1}) = 0, \qquad \psi_n(\alpha_n) = 1.$$

Then the sequence $\{\psi_n(z)\}$ is orthogonal on C.

Theorem 2 is similar to a theorem due to Bergman [3], involving analytic functions and multiple interpolation in a single point of a region.

The set of functions of class H satisfying the auxiliary conditions (13) is not empty, for there exists a polynomial in z satisfying those conditions; the real part of this polynomial is of class H and also satisfies those conditions. Thus $\psi_n(z)$ exists. No function of class H satisfying (13) is identically zero.

The function $\psi_n(z)$ defined in Theorem 2 is orthogonal to a function $u(z)$ of class H with $u(\alpha_0) = u(\alpha_1) = \cdots = u(\alpha_{n-1}) = 0$ when and only when we have $u(\alpha_n) = 0$; we omit the proof, which follows precisely the method of proof of Lemma 1. Consequently each ψ_n is orthogonal to $\psi_{n+1}, \psi_{n+2}, \cdots$, so the sequence is orthogonal.

Moreover, if we have for a function $u(z)$ of class H the equations $u(a_0) = u(\alpha_1) = \cdots = u(\alpha_{n-1}) = 0$, then we have as in the proof of (6),

$$(14) \qquad u(\alpha_n) = \int_C u(z)\psi_n(z) \mid dz \mid / \int_C \psi_n(z) \mid dz \mid.$$

A special case of this equation is found by setting $u(z) \equiv \psi_n(z)$:

$$1 = \int_C [\psi_n(z)]^2 \mid dz \mid / \int_C \psi_n(z) \mid dz \mid,$$

so (14) can be written

$$(15) \qquad u(\alpha_n) = \int_C u(z)\psi_n(z) \mid dz \mid / \pi \parallel \psi_n \parallel^2.$$

If $u(z)$ is an arbitrary function of class H, there are two formal expansions of $u(z)$ of form

$$(16) \qquad u(z) \sim a_0\psi_0(z) + a_1\psi_1(z) + \cdots;$$

the one expansion is the usual orthogonal function (Fourier type) expansion found by formal expansion of $u(z)$ on C, and the other is an interpolation series expansion found by setting successively $z = \alpha_0, \alpha_1, \cdots$, and solving for $a_0, a_1, \cdots$ in order:

$$u(\alpha_0) = a_0\psi_0(\alpha_0), \qquad u(\alpha_1) = a_0\psi_0(\alpha_1) + a_1\psi_1(\alpha_1), \cdots.$$

These two formal expansions are identical. The fact that a_0 is the same

whether found by one method or the other is a consequence of (15) for $n = 0$. Assume this conclusion true for $a_0, a_1, \cdots, a_{n-1}$. The function

$$u_{n-1}(z) \equiv u(z) - a_0\psi_0(z) - a_1\psi_1(z) - \cdots - a_{n-1}\psi_{n-1}(z)$$

vanishes for $z = \alpha_0, \alpha_1, \cdots, \alpha_{n-1}$. By (15) it is then immaterial whether a_n is determined by interpolation to $u_{n-1}(z)$ in $z = \alpha_n$, namely $a_n\psi_n(\alpha_n) = u_{n-1}(\alpha_n)$, with $\psi_n(\alpha_n) = 1$, or determined by the formal orthogonal expansion of $u_{n-1}(z)$ on C: $\pi a_n \| \psi_n \|^2 = \int_C u_{n-1}(z)\psi_n(z) | dz |$.

Each function $\psi_n(z)$ is unique, and is also the unique function which satisfies the auxiliary conditions (13) and is orthogonal to every $u(z)$ of class H which satisfies the conditions $u(\alpha_0) = \cdots = u(\alpha_n) = 0$, as follows by the method already used in connection with Theorem 1 for $n = 0$. We shall prove that $\psi_n(z)$ *is a linear combination of the functions* $\Psi_0(z)$, $\Psi_1(z), \cdots, \Psi_n(z)$, where $\Psi_k(z) \equiv f_{\alpha_k}(z)$. These functions $\Psi_k(z)$ are linearly independent on C, for by (12) they are functions harmonic over the extended plane except for singularities in the distinct points $1/\bar{\alpha}_k$ respectively. These functions $\Psi_k(z)$ can then be orthogonalized and normalized on C by the Gram-Schmidt process, yielding a set $\phi_0(z), \phi_1(z), \cdots, \phi_n(z)$, where $\phi_k(z)$ is a linear combination of the functions $\Psi_0(z), \Psi_1(z), \cdots, \Psi_k(z)$. Thus $\phi_k(z)$ is orthogonal on C to the functions $\phi_0, \phi_1, \cdots, \phi_{k-1}$, hence is orthogonal to $\Psi_0, \Psi_1, \cdots, \Psi_{k-1}$, hence by Theorem 1 vanishes in the points $\alpha_0, \alpha_1, \cdots, \alpha_{k-1}$. But ϕ_k is not orthogonal to Ψ_k, so (also by Theorem 1) we have $\phi_k(\alpha_k) \neq 0$. Every function $u(z)$ which vanishes in all the points $\alpha_0, \alpha_1, \cdots, \alpha_k$ is orthogonal to $\Psi_0, \Psi_1, \cdots, \Psi_k$, hence is orthogonal to $\phi_k(z)$. By our characterization of $\psi_k(z)$ in terms of orthogonality to such functions $u(z)$, it follows that $\phi_k(z)$ is a constant multiple of $\psi_k(z)$, so $\psi_k(z)$ is a linear combination of the functions $\Psi_0, \Psi_1, \cdots, \Psi_k$; conversely, Ψ_k is a linear combination of $\psi_0, \psi_1, \cdots, \psi_k$.

The formulas for the orthogonalization and normalization of the functions (12) may be readily written by the reader in determinantal form; we have for instance

$$\int_C \Psi_j(z)\Psi_k(z) | dz | = \int_C \Psi_j(z) f_{\alpha_k}(z) | dz | = \Psi_j(\alpha_k) \int_C f_{\alpha_k}(z) | dz |$$
$$= 2\pi[(1 - |\alpha_j|^2)/(1 + |\alpha_j|^2)][(1 - |\alpha_k|^2)/(1 + |\alpha_k|^2)]$$
$$\times \mathrm{Re}[(1 + \bar{\alpha}_j\alpha_k)/(1 - \bar{\alpha}_j\alpha_k)].$$

The two fundamental questions of interpolation for functions of class H are 1) given a function $u(z)$ of class H and a set of points $\alpha_0, \alpha_1, \cdots$ interior to C; to express $u(z)$ in terms of the values $u(\alpha_k)$; 2) given a set

of points $\alpha_0, \alpha_1, \cdots$ interior to C and corresponding functional values A_0, $A_1, \cdots$; to determine all functions $u(z)$ of class H satisfying $u(\alpha_k) = A_k$. The second of these questions is the more general and essentially includes the first, so we proceed to consider it.

There obviously exists a unique linear combination of the functions $\psi_0(z), \psi_1(z), \cdots, \psi_n(z)$ which takes on prescribed values $A_0, A_1, \cdots, A_n$ in the respective points $\alpha_0, \alpha_1, \cdots, \alpha_n$, so there exists a unique linear combination of the functions $\Psi_0(z), \Psi_1(z), \cdots, \Psi_n(z)$ which takes on those values in the respective points. The formula for these (equivalent) linear combinations is readily written in determinantal form. The function $u_0(z)$ thus obtained is orthogonal to every function $u(z)$ of class H which vanishes in the points $\alpha_0, \alpha_1, \cdots, \alpha_n$, hence is the unique function of minimum norm which satisfies the prescribed conditions of interpolation:

$$\| u_0 + u \|^2 = \| u_0 \|^2 + \| u \|^2 \geqq \| u_0 \|^2.$$

The case of an infinite number of points $\alpha_0, \alpha_1, \cdots$ and prescribed values $A_0, A_1, \cdots$ is similarly treated. The points and values define a formal infinite expansion (16) found by interpolation in the points α_k; a necessary and sufficient condition that there exist a function $u(z)$ of class H taking on the values A_k in the points α_k is the convergence of the series $\sum a_k^2 \| \psi_k \|^2$. If this condition is fulfilled, the second member of (16) converges in the mean on C, hence by (7) converges uniformly on any $| z | \leqq r < 1$, and represents a function $u_0(z)$ in H which satisfies the conditions of interpolation. Any function in H which vanishes in all the points α_k is orthogonal to every ψ_k, and conversely; such a function is orthogonal to $u_0(z)$. Thus $u_0(z)$ is the unique function of class H of least norm which satisfies the conditions of interpolation. The most general function of class H satisfying the conditions of interpolation is then $u_0(z)$ plus an arbitrary function $u_1(z)$ of class H which vanishes in all the points α_k, or otherwise expressed is $u_0(z)$ plus an arbitrary function $u_1(z)$ of class H which is orthogonal to all the $\psi_k(z)$.

If we take as departure not arbitrary assigned values A_k but a function $u(z)$ of H, its formal series of interpolation (16) is on C merely the formal expansion of $u(e^{i\theta})$ (measurable and square-integrable on C) in terms of the orthogonal functions $\psi_k(e^{i\theta})$, hence converges in the mean on C and represents in $| z | < 1$ the function $u_0(z)$ of least norm satisfying the conditions $u_0(\alpha_k) = u(\alpha_k)$ for every k.

3. Real points of interpolation. There is one non-trivial case where the functions $\phi_k(z)$ and $\psi_k(z)$ can be written down at once:

THEOREM 3. *Let there be given the real distinct points*

$$\alpha_0 = 0, \alpha_1, \alpha_2, \cdots \tag{17}$$

interior to C. In series (16) *we may choose* $\psi_0(z) \equiv 1$,

$$\psi_n(z) \equiv \mathrm{Re}[\lambda_n z(z-\alpha_1)\cdots(z-\alpha_{n-1})/\{(1-\alpha_1 z)\cdots(1-\alpha_n z)\}],$$

$$\lambda_n = (1-\alpha_1\alpha_n)\cdots(1-\alpha_n^2)/\alpha_n(\alpha_n-\alpha_1)\cdots(\alpha_n-\alpha_{n-1}). \tag{18}$$

The functions

$$2^{-\frac{1}{2}},\ \mathrm{Re}[(1-|\alpha_n|^2)^{\frac{1}{2}} z(z-\alpha_1)\cdots(z-\alpha_{n-1})/\{(1-\alpha_1 z)\cdots(1-\alpha_n z)\}]$$

are normal and orthogonal on C.

The analytic functions $F_n(z) \equiv U_n(z) + iV_n(z)$ whose real parts occur in (18) are known ([1], p. 305) to be orthogonal on C; for $j \neq k$ we have

$$0 = \int_C F_j(z)\bar{F}_k(z)|\,dz\,| = \int_C (U_j + iV_j)(U_k - iV_k)\,ds,$$

$$\int_C (U_jU_k + V_jV_k)\,ds = \int_C (U_jV_k - U_kV_j)\,ds = 0. \tag{19}$$

On C we have $dz = iz\,ds$, so by Cauchy's integral formula (even if $j = k \neq 0$)

$$\int_C F_j(z)F_k(z)\,ds = -i\int_C F_j(z)F_k(z)\,dz/z = 0,$$

$$\int_C (U_jU_k - V_jV_k)\,ds = \int_C (U_jV_k + U_kV_j)\,ds = 0. \tag{20}$$

The orthogonality of the set exhibited in (18) follows from (19) and (20).

The function $\psi_n(z)$ exhibited in (18) is a linear combination *with real coefficients* of the functions

$$\mathrm{Re}[(1+\alpha_k z)/(1-\alpha_k z)], \qquad k = 0, 1, 2, \cdots, n, \tag{21}$$

by the classical theory of the decomposition of rational functions; since the procedure of orthogonalization of a given sequence leads to essentially unique results, it follows that the sequence exhibited in (18), found by orthogonalization of the sequence (21), is identical with the sequence $\psi_n(z)$ previously considered in (16), found by orthogonalization of the functions $\Psi_k(z) \equiv f_{\alpha_k}(z)$ of (12).

From the analogue of (19) with $j = k > 0$ and from (20) we have $\int_C |F_j|^2\,ds = 2\int_C U_j^2\,ds$; the first member is readily computed, so normalization is as indicated in Theorem 3.

4. Completeness of orthogonal functions. The functions $\psi_k(z)$ in (16) may be complete but need not be complete. For instance, if the α_k can be written in the form $\alpha_k = \rho e^{ik\beta}$, $0 < \rho < 1$, where ρ is independent of k and where β/π is irrational, the points α_k are everywhere dense on the circle $|z| = \rho$; a function of class H which vanishes in all the points α_k vanishes identically interior to C. On the other hand, if the α_k satisfy the conditions of Theorem 3, the function $V_k(z)$, $k \neq 0$, in H does not vanish identically and is orthogonal to all the functions $\psi_n(z)$ [or $U_n(z)$], so the latter do not form a complete set in H.

There are various procedures for completing a given set $\psi_k(z)$ in (16), to which we now turn. An arbitrary function of class H which vanishes in all the points α_k is orthogonal to each $\psi_k(z)$, and conversely. If the functional values $A_k = u(\alpha_k)$ are prescribed with $\sum a_k^2 \| \psi_k \|^2$ convergent, or if a function $u(z)$ of class H is given, the formal expansion (16) converges in $|z| < 1$ and converges in the mean on C to the function $u_0(z)$ of minimum norm with the property $u_0(\alpha_k) = u(\alpha_k)$. Every function $u(z)$ of class H which takes on the prescribed values is given by the equation

$$u(z) \equiv u_0(z) + u_1(z), \tag{22}$$

where $u_1(z)$ is a function of class H which vanishes in all the points α_k. By the orthogonality of u_0 and u_1 we have

$$\| u \|^2 = \| u_0 \|^2 + \| u_1 \|^2. \tag{23}$$

If the set $\psi_k(z)$ in (16) is not complete, we denote by H' the closure of the linear family of functions $u(z)$ of H defined by the $\psi_k(z)$, and denote by H'' the closure of the linear family defined by a complete subset of orthogonal functions of H each of which is orthogonal to all the ψ_n; then H'' is the class of all functions of H vanishing in all the points α_k, or alternately orthogonal to all the ψ_n. Each function of H is the unique sum of a function of H' and a function of H''. It may occur (in contrast to the facts for interpolation to analytic functions, for which the set of points on which all functions of the analogue of H'' vanish is precisely the set $\{\alpha_k\}$, and the Blaschke product whose zeros are the α_k is convergent) that *each function of H'' vanishes on a point set larger than* $\{\alpha_k\}$, as for instance under the conditions of Theorem 3 with $\alpha_k \to 0$; in the notation of the proof of Theorem 3, the functions $F_n(z) \equiv U_n(z) + iV_n(z)$ form a complete set with respect to the class H_2 of analytic functions $\sum c_k z^k$, $\sum |c_k|^2 < \infty$; the real part of a function of class H_2 is of class H, and an analytic function whose real part is of class H is itself of class H_2; thus the functions $\{U_n, V_n\}$ form

a complete set with respect to class H; the closure of the linear family defined by $\{U_n\}$ is H', and the closure of the linear family defined by $\{V_n\}$ is H''; all functions of the latter family vanish at every point of the axis of reals interior to C.

We can be more explicit, with the choice that α_k is real, $\alpha_k \to 0$. Every function of class H which vanishes for real z is orthogonal to each $\psi_k(z)$, hence is orthogonal to each function of class H'. But it follows from (2) that every function of class H is an analytic function of the real variable r on the segment $-1 < x < +1$, $y = 0$, so every function of class H which is orthogonal to all the $\psi_k(z)$ vanishes in all the points α_k and vanishes on the axis of reals, hence (by Schwarz's principle of symmetry) satisfies the equation $u(z) \equiv -u(\bar{z})$. In the expansion (2), which is unique, it follows that every a_k vanishes. Consequently, the class H'' is precisely the class of functions (2) with every a_k equal to zero, namely the closed linear extension of the class $\{r^k \sin k\theta\}$; otherwise expressed, the class H'' is the subclass of H for which $u(e^{i\theta})$ is an odd function of θ.

We return to the situation of arbitrary α_k where the set H'' is not empty, and denote by S the set of points interior to C on which not all the functions of H'' vanish. Precisely the interpolation method used in the proof of Theorem 1 yields a formula analogous to (6) for the determination of the value in an arbitrary point α of S of an arbitrary function of class H''. Moreover there can be determined by the methods already used a set of orthogonal functions $\psi''_0, \psi''_1, \cdots$ all of class H'' such that an arbitrary function $u''(z)$ of class H'' can be represented

$$u''(z) \sim a_0\psi''_0 + a_1\psi''_1(z) + \cdots; \tag{24}$$

the function $\psi''_k(z)$ vanishes in points $\alpha''_0, \alpha''_1, \cdots, \alpha''_{k-1}$ of S and does not vanish in α''_k; the unique formal expansion (24) may be found either by the usual method of expansion on C in orthogonal functions or alternately by formal interpolation to $u''(z)$ in the points α''_k; the points α''_k are not uniquely determined, but may be chosen for instance as the subset in S of the points $\rho e^{ik\beta}$, $0 < \rho < 1$, $k = 0, 1, 2, \cdots$, where β/π is irrational, and ρ is independent of k. The functions ψ''_k are all orthogonal to the original set ψ_k, so an arbitrary function $u(z)$ of class H has the expansion in orthogonal functions

$$u(z) \sim a_0\psi_0 + a_1\psi_1 + \cdots + a''_0\psi''_0 + a''_1\psi''_1 + \cdots, \tag{25}$$

which is valid in $|z| < 1$ and valid in the mean on C. Series (25) also partakes of the usual properties of interpolation series; the function ψ_k

vanishes in the points $\alpha_0, \alpha_1, \cdots, \alpha_{k-1}$ but not in the point α_k, so if $a_0, a_1, \cdots, a_{k-1}$ have been determined, the coefficient a_k can be determined by formal interpolation in the point α_k; the function ψ''_k vanishes in the points $\alpha_0, \alpha_1, \cdots, \alpha''_0, \alpha''_1, \cdots, \alpha''_{k-1}$ but not in the point α''_k, so if $a_0, a_1, \cdots, a''_0, a''_1, \cdots, a''_{k-1}$ have been determined, the coefficient a''_k can be determined by formal interpolation in the point α''_k.

5. Kernel functions. The relation of the classes H, H', H'' is of particularly simple form when expressed in terms of kernel functions. If the set of functions $\phi_n(z)$ each of class H is normal and orthogonal but not necessarily complete on C, and if an arbitrary sequence a_n with $\sum a_n^2$ convergent is chosen, the series $\sum_0^\infty a_n\phi_n(z)$ converges in the mean on C, and hence by (7) converges uniformly on every closed set interior to C. Since the series $\sum a_n\phi_n(z)$ converges for every such choice of the a_n, it follows by a result due to Landau that the series $\sum[\phi_n(z)]^2$ converges for every z interior to C. Consequently the *kernel function*

$$k(z,t) \equiv \frac{1}{\pi}\sum_{n=0}^{\infty}\phi_n(z)\phi_n(t) \tag{26}$$

is defined for each t with $|t|<1$ as a function of class H in z, and is symmetric in z and t; the series defining $k(z,t)$ converges in the mean for z on C. If $u(z)$ is a function of H and if the set $\phi_n(z)$ is complete in H, we may write for every z with $|z|<1$

$$\int_C k(z,t)u(t)\,|\,dt\,| \equiv \frac{1}{\pi}\sum_{n=0}^{\infty}\phi_n(z)\int_C u(t)\phi_n(t)\,|\,dt\,| \equiv u(z).$$

From this equation it follows that the kernel when defined by (26) in terms of a complete set $\{\phi_n\}$ of normal and orthogonal functions can also be defined by the analogue of (26) in terms of a second complete set $\{\phi'_n\}$ of normal orthogonal functions:

$$\int_C k(z,t)\phi'_n(t)\,dt \equiv \phi'_n(z), \qquad k(z,t) \equiv \frac{1}{\pi}\sum_{n=0}^{\infty}\phi'_n(z)\phi'_n(t),$$

for $|t|<1$, since $k(z,t)$ is of class H in z for fixed t, $|t|<1$; this series for $k(z,t)$ converges in the mean for z on C and uniformly for $|z|\leqq r<1$.

A kernel k, k', k'' is associated with each of the classes H, H', H'' previously introduced; the kernel can be expressed in terms of a complete normal orthogonal set for the corresponding class:

$$k(z,t) \equiv \tfrac{1}{\pi} \sum \phi_n(z)\phi_n(t), \qquad k'(z,t) \equiv \tfrac{1}{\pi} \sum \psi'_n(z)\psi'_n(t),$$
$$k''(z,t) \equiv \tfrac{1}{\pi} \sum \psi''_n(z)\psi''_n(t).$$

Since the set $\{\psi'_n, \psi''_n\}$ is complete in H, we have $k(z,t) \equiv k'(z,t) + k''(z,t)$. If $u(z)$ is of class H, we have

$$\int_C k'(z,t)u(t)\,|\,dt\,| \equiv \tfrac{1}{\pi} \sum_{n=0}^{\infty} \psi'_n(z) \int_C u(t)\psi'_n(t)\,|\,dt\,|,$$

which is the formal expansion of $u(z)$ in terms of the set $\{\psi'_n\}$.

If a function $u(z)$ of class H is given, or if functional values in the points α_k are otherwise given corresponding to which a function of class H exists, we have a formal development $\sum a_k\psi'_k(z)$; if $u(z)$ is given this development represents the function $\int_C k'(z,t)u(t)\,|\,dt\,|$. If the functions $\psi'_k(z)$ are complete in H, this development represents the unique function of class H taking on the given values in the given points. If the functions $\{\psi'_k(z)\}$ are not complete in H, this development represents the function of class H of least norm taking on the given values in the given points; the whole class of functions of H taking on the given values in the points α_k is precisely the class

$$\sum_{k=0}^{\infty} a_k\psi'_k(z) + \sum_{k=0}^{\infty} b_k\psi''_k(z), \text{ where } \sum b_k^2 < \infty,$$

and can also be expressed as $\sum_{k=0}^{\infty} a_k\psi'_k(z) + \int_C k''(z,t)U(t)\,|\,dt\,|$, where $U(z)$ is an arbitrary function of class H.

In the special case already considered (§ 4) that the α_k are real with $\alpha_k \to 0$, we have shown that H'' is precisely the set of functions $u(z)$ in H for which $u(e^{i\theta})$ is an odd function of θ; since every function of θ is the unique sum of an odd function of θ and an even function of θ, and since every odd function is orthogonal to every even function, it follows that the set H' of functions of H orthogonal to all functions of H'' is precisely the set of functions $u(z)$ of H for which $u(e^{i\theta})$ is an even function of θ. For this special case we may write $(r < 1, \rho < 1)$

$$k'(re^{i\theta}, \rho e^{i\phi}) \equiv \tfrac{1}{2\pi} + \tfrac{1}{\pi} \sum_{n=1}^{\infty} r^n\rho^n \cos n\theta \cos n\phi$$
$$\equiv \tfrac{1}{2\pi} \operatorname{Re}[1/(1 - r\rho e^{i(\theta+\phi)}) + r\rho e^{i(\theta-\phi)}/(1 - r\rho e^{i(\theta-\phi)})],$$

$$k''(re^{i\theta}, \rho e^{i\phi}) \equiv \tfrac{1}{\pi} \sum_{n=1}^{\infty} r^n\rho^n \sin n\theta \cos n\phi$$
$$\equiv \tfrac{1}{2\pi} \operatorname{Re}[r\rho e^{i(\theta-\phi)}/(1 - r\rho e^{i(\theta-\phi)}) - r\rho e^{i(\theta+\phi)}/(1 - r\rho e^{i(\theta+\phi)})].$$

6. Multiple points. In interpolation problems for analytic functions, coincidence of two points of interpolation corresponds to prescription of the derivative of a function; for harmonic functions, such coincidence corresponds to prescription of a partial derivative of a function; similar remarks apply to coincidence of more than two points of interpolation. For harmonic functions it is obvious by Laplace's equation that not all partial derivatives of a function at a point can be assigned arbitrarily; indeed, at any point precisely two partial derivatives of any given order can be assigned independently. For instance, a harmonic function $u(z)$ which vanishes at $z=0$ together with its $(n-1)$-st derivatives but with $\partial^n u(0)/\partial x^n \neq 0$ is

$$u(z) \equiv \mathrm{Re}(z^n) \equiv x^n - n(n-1)x^{n-2}y^2/2 + \cdots;$$

a harmonic function $v(z)$ which vanishes at $z=0$ together with its $(n-1)$-st derivatives but with $\partial^n v(0)/\partial x^{n-1}\partial y \neq 0$ is $v(z) \equiv \mathrm{Im}(z^n) \equiv nx^{n-1}y - \cdots$. The expansion (2) can be interpreted as an interpolation series defined from the values of $u(z)$ and the derivatives $\partial^n u/\partial x^n$ and $\partial u^n/\partial x^{n-1}\partial y$ at $z=0$; the corresponding complete set of orthogonal functions of class H is $\{r^n \cos n\theta, r^n \sin n\theta\}$.

The theory of interpolation for harmonic functions as we have already developed it, holds with only minor and obvious changes if partial derivatives as well as functional values are assigned. For instance if $z_0(|z_0|<1)$ and ν are given, there exists a function $u_0(z)$ in H with $\partial^\nu u_0(z_0)/\partial x^\nu = 1$ whose norm is least; any function $u(z)$ in H satisfies the equation $\partial^\nu u(z_0)/\partial x^\nu = 0$ when and only when $u(z)$ and $u_0(z)$ are orthogonal; we have for an arbitrary function $u(z)$ of H

$$\int_C u_0(z)[u(z) - u_0(z)\partial^\nu u(z_0)/\partial x^\nu]ds = 0,$$

$$\partial^\nu u(z_0)/\partial x^\nu = \int_C u_0(z)u(z)ds \Big/ \int_C [u_0(z)]^2 ds,$$

a formula which corresponds to differentiation of Poisson's integral under the integral sign. The choice $u(z) \equiv U_0(z)$, a multiple of $\mathrm{Re}[(z-z_0)^n]$ with $\partial^\nu U_0(z_0)/\partial x^\nu = 1$, yields $\int_C [u_0(z)]^2 ds = \int_C u_0(z)U_0(z)ds$. We shall not attempt to derive specific formulas to include all such cases. The set of functions (18) with the α_k not all distinct corresponds to prescription of every $u(\alpha_k)$ and of suitable successive derivatives $\partial^n u(\alpha_k)/\partial x^n$ in multiple points α_k.

If a series of interpolation for harmonic functions is desired, corre-

sponding to interpolation in points some of which are counted multiply, it is necessary to order the points according to increasing multiplicities, so that functions $\psi_n(z)$ of increasing subscript correspond to derivatives of higher order at a point.

We have chosen (3) as the basis of our definition of norm because of its close relation to Poisson's integral and the Dirichlet problem and to other classical results. Several generalizations of our entire discussion can be envisaged: a) use of a positive norm function in (3), b) use of surface integrals over $|z| < 1$ in (1) and (3) instead of line integrals, and c) consideration of more general regions, not necessarily simply connected. The methods that we have used apply also in these cases; the specific formulas are somewhat different, but at least in the simpler cases can be obtained by the reader.

HARVARD UNIVERSITY.

REFERENCES.

[1] J. L. Walsh, "Interpolation and approximation," *Colloquium Publications of the American Mathematical Society*, vol. XX, New York, 1935.

[2] O. Lokki, "Über analytische Funktionen, deren Dirichlet-integral endlich ist und die in gegebenen Punkten vorgeschriebene Werte annehmen," *Annales Academiae Scientiarum Fennicæ*, AI39 (1947).

[3] S. Bergman, "The kernel function and conformal mapping," *Mathematical Surveys*, vol. V (1950), New York.

[4] J. L. Walsh and Philip Davis, "Interpolation and orthonormal systems," *Journal d'Analyse Mathématique*, vol. 2 (1952), pp. 1-28.

OHIO UNIVERSITY
LIBRARY

SUR L'HOMOLOGIE ET LA COHOMOLOGIE DES GROUPES DE LIE COMPACTS CONNEXES.*

par ARMAND BOREL.

Introduction. On sait que l'homologie et la cohomologie à coefficients réels des groupes de Lie simples compacts connexes sont complètement connues, (*voir* notamment [7], [10], [16], [17], [18]†). Quant à la cohomologie entière ou mod. p, elle a été déterminée pour les groupes classiques seulement, tant au point de vue additif ([8], [9], [16]) que multiplicatif ou même des puissances réduites de Steenrod ([3], [4], [6], [15]).[1] Nous nous proposons ici principalement d'étudier la cohomologie par rapport à un corps de caractéristique quelconque des groupes localement isomorphes aux groupes classiques et des deux premiers groupes exceptionnels $\boldsymbol{G}_2$, $\boldsymbol{F}_4$; pour $\boldsymbol{Spin}(n)$, $\boldsymbol{G}_2$, $\boldsymbol{F}_4$ nous déterminerons aussi la cohomologie entière et, partiellement, l'algèbre d'homologie (le produit étant le produit de Pontrjagin), et les éléments universellement transgressifs.

Ces résultats feront l'objet des 3ème, 4ème, 5ème parties, qui sont précédées de généralités sur le produit de Pontrjagin et d'un théorème de transgression dans les espaces universels; ils seront établis essentiellement par l'étude de l'algèbre spectrale de certaines fibrations, les unes de nature générale faisant intervenir les espaces classifiants, les autres particulières aux groupes étudiés, et en grande partie indépendamment des résultats de I et II: Les Nos. 3 à 6 notamment ne seront utilisés que dans les Nos. 15 et 16. Pour pouvoir décrire ces algèbres spectrales et, éventuellement, passer à l'algèbre dont E_∞ est l'algèbre graduée, nous nous appuyerons fréquemment sur des résultats de [3], pour la plupart longuement résumés dans [6], et qui ici seront supposés connus. Toutefois nous avons rappelé au No. 0 les principales notations et au No. 1, avec quelques compléments, toutes les définitions et résultats concernant la structure de l'algèbre de cohomologie d'un H-espace.

---

* Received October 7, 1953.

† ainsi qu'une Note écrite en collaboration avec C. Chevalley, à paraître dans les Memoirs of the Amer. Math. Soc.

[1] Ces résultats sont rappelés au No 8.

Le No. 2 établit quelques relations entre le produit de Pontrjagin et l'algèbre de cohomologie; le No. 3 associe à tout élément $x \in H_s(G, K_p)$ un endomorphisme diminuant le degré de s de l'espace de cohomologie $H^*(G, K_p)$, ou encore de la suite spectrale d'un espace fibré principal de groupe structural G; ces opérateurs ont été introduits en caractéristique 0 par J. Leray [12], [13]; comme application, nous démontrons un théorème assez général contenant comme cas particuliers des résultats de J. Leray, H. Samelson relatifs aux homomorphismes transposés des applications $G \to G/U$, $U \to G$, (U sous-groupe de G).

Dans II nous montrons que si l'algèbre de cohomologie $H^*(B_G, K_p)$ d'un espace classifiant pour G est une algèbre de polynômes, alors $H^*(G, K_p)$ possède un système simple de générateurs universellement transgressifs; c'est donc, grosso modo, une réciproque aux résultats du § 19 de [3].

On sait que les centres des groupes classiques $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$, $\boldsymbol{SO}(2n)$, $\boldsymbol{SO}(2n+1)$,[2] sont cycliques à respectivement n, 2, 2, 1 éléments. Dans la 3ème partie, nous nous occupons des quotients de ces groupes par leurs centres (ou dans le premier cas par un sous-groupe du centre), autrement dit des groupes projectifs unitaires $\boldsymbol{PU}(n)$, $\boldsymbol{PSp}(n)$, $\boldsymbol{PSO}(2n)$. Auparavant, nous étudions plus généralement au No. 10 la cohomologie mod. p du quotient de G par un sous-groupe cyclique Z_m, ($m \equiv 0 \bmod p$), à l'aide de l'algèbre spectrale des revêtements réguliers, sur laquelle on peut donner des renseignements très précis lorsque $H^*(G, K_p)$ possède un système simple de générateurs universellement transgressifs.

IV est consacrée au revêtement universel, à deux feuillets, du groupe orthogonal $\boldsymbol{SO}(n)$, le groupe des spineurs $\boldsymbol{Spin}(n)$. L'essentiel est ici la cohomologie mod 2 et nous calculons $H^*(\boldsymbol{Spin}(n), Z_2)$ au No. 13, par l'intermédiaire de l'algèbre spectrale de la fibration $\boldsymbol{Spin}(n)/\boldsymbol{S}_1 = \boldsymbol{V}_{n,n-2}$; en particulier, $\boldsymbol{Spin}(n)$ a de la 2-torsion si et seulement si $n \geqq 7$, et son polynôme de Poincaré mod 2 s'obtient à partir de $(1+t)(1+t^2)\cdots(1+t^{n-1})$ en y remplacant $(1+t^{k-1})\cdot(1+t^k)$ par $(1+t^{2k-1})$ toutes les fois que k est une puissance de deux; le No. 14 montrera que les coefficients de torsion de $\boldsymbol{Spin}(n)$, ($n \geqq 7$), sont tous égaux à deux. Nous étudions ensuite aux Nos. 15, 16 les éléments universellement transgressifs et le produit de Pontrjagin mod 2 de $\boldsymbol{Spin}(n)$; malgré les calculs assez longs et fastidieux auxquels on est ainsi conduit, il nous a semblé qu'il valait la peine d'insister sur ces points où le comportement

---

[2] *Voir le* No. 0 pour les notations.

de $\boldsymbol{Spin}(n)$, $(n \geqq 10)$, paraît quelque peu exceptionnel; en effet, tandis que pour $G = \boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$, $\boldsymbol{SO}(n)$, $\boldsymbol{G}_2$, $\boldsymbol{F}_4$, l'algèbre $H^*(G, Z_2)$ est engendrée par des éléments universellement transgressifs, et que $H_*(G, Z_2)$ est commutative, on constatera que $\boldsymbol{Spin}(n)$ ne possède aucune de ces deux propriétés pour $n \geqq 10$.

L'étude des deux premiers groupes exceptionnels est faite dans V. Celle de $\boldsymbol{G}_2$, très simple, utilise les deux fibrations $\boldsymbol{G}_2/\boldsymbol{S}_3 = \boldsymbol{V}_{7,2}$ et $\boldsymbol{Spin}(7)/\boldsymbol{G}_2 = \boldsymbol{S}_7$; on trouve notamment que les coefficients de torsion de $\boldsymbol{G}_2$ sont tous égaux à deux. Pour $\boldsymbol{F}_4$ nous nous servirons des fibrations déduites des inclusions $\boldsymbol{F}_4 \supset \boldsymbol{Spin}(9) \supset \boldsymbol{Spin}(8) \supset \boldsymbol{T}^4$ et $\boldsymbol{F}_4 \supset \boldsymbol{Spin}(9) \supset \boldsymbol{Spin}(7) \supset \boldsymbol{G}_2$; un point important est le fait que le groupe des permutations de trois objets opère fidèlement sur $H^*(\boldsymbol{F}_4/\boldsymbol{Spin}(8), Z)$, trivialement sur $H^*(\boldsymbol{F}_4, Z)$ et commute avec l'homomorphisme naturel du premier groupe dans le second; c'est ce qui nous permettra de voir que $\boldsymbol{F}_4$ n'a pas de p-torsion pour $p > 3$. Par contre les valeurs de $H^*(\boldsymbol{F}_4, Z_3)$ et de $H^*(\boldsymbol{F}_4, Z_2)$, (Nos. 21-22), montreront que $\boldsymbol{F}_4$ a de la 2- et de la 3-torsion; on vérifiera enfin (No. 23) que ses coefficients de torsion sont égaux à 2, 3 ou 6.

Pour conclure cette vue d'ensemble des résultats acquis sur les groupes de Lie simples compacts connexes, signalons que l'on ne sait rien sur la torsion des trois derniers groupes exceptionnels, mis à part le fait que $\boldsymbol{E}_8$ a de la 2-torsion [5].

Les principaux résultats de ce travail ont été résumés dans une Note aux Proc. Nat. Acad. Sci. U. S. A. *39* (1953), pp. 1142-1146. Une partie d'entre eux avait été antérieurement annoncée aux *Comptes Rendus* [2].

TABLE DES MATIERES

Introduction.

No. 0. Préliminaires; principales notations.

I. Le Produit de Pontrjagin

No. 1. Cohomologie d'un H-espace.
No. 2. Le produit de Pontrjagin.
No. 3. Opérateurs sur la cohomologie d'un espace fibré principal définis par l'homologie du groupe structural.

II. Un Théorème de Transgression

No. 4. Énoncé d'une proposition et notations.
No. 5. Démonstration de la Proposition 4. 1.
No. 6. Un théorème de transgression.

III. Groupes Quotients des Groupes Classiques

No. 7. Sur l'algèbre spectrale des espaces fibrés.
No. 8. Homologie et cohomologie des groupes classiques.
No. 9. Revêtements réguliers finis, cohomologie des groupes cycliques.
No. 10. Quotient d'un groupe de Lie par un sous-groupe cyclique.
No. 11. Groupes quotients des groupes classiques.

IV. Le Groupe des Spineurs

No. 12. Cohomologie du groupe des spineurs (résultats).
No. 13. Cohomologie mod 2.
No. 14. Cohomologie entière.
No. 15. Éléments universellement transgressifs de $H^*(\boldsymbol{Spin}(n), Z_2)$.
No. 16. Sur l'homologie du groupe des spineurs.

V. Les Deux Premiers Groupes Exceptionnels

No. 17. Le groupe $\boldsymbol{G}_2$.
No. 18. Groupes opérant sur un espace fibré principal.
No. 19. Homologie et cohomologie de $\boldsymbol{F}_4$, (résultats).
No. 20. L'homomorphisme $H^*(\boldsymbol{F}_4/\boldsymbol{Spin}(8), Z) \to H^*(\boldsymbol{F}_4, Z)$.
No. 21. Cohomologie mod p de $\boldsymbol{F}_4$, $(p \neq 2)$.
No. 22. Cohomologie mod 2 de $\boldsymbol{F}_4$.
No. 23. Cohomologie entière de $\boldsymbol{F}_4$.

0. Préliminaires; principales notations.

0. 1. Algèbre. p désignera toujours soit un nombre premier soit zéro, K_p un corps de caractéristique p, Z_p, $(p \neq 0)$, le corps des entiers mod. p, Z_0 celui des rationnels.

Tors. A est le sous-groupe des éléments d'ordre fini du groupe abélien A, Tors. ${}_pA$ sa composante p-primaire; (U) dénotera l'idéal bilatère engendré par une partie U d'une algèbre V.

L'algèbre extérieure d'un espace vectoriel sur un corps que le contexte précisera (ou d'un groupe abélien libre) P est notée $\wedge P$ ou encore $\wedge(x_1, \cdots, x_m)$, si (x_i) est une base de P, et $\wedge^j P$ est bien entendu le sous-espace (sous-groupe) engendré par les produits extérieurs de j éléments de P.

Soient H une algèbre unitaire, $x_1, \cdots, x_m$ des éléments de H, et P le sous-espace qu'ils engendrent. On notera $\triangle(x_1, \cdots, x_m)$ ou $\triangle P$ le sous-espace engendré par 1 et par les monômes $x_{i_1} \cdots x_{i_k}$, $(i_1 < \cdots < i_k; 1 \leqq k \leqq m)$, lorsque ces monômes sont linéairement indépendants, (*voir* la définition de système simple de générateurs au No. 1).

0. 2. Homologie et cohomologie. $H_i(X, \Gamma)$, (resp. $H^i(X, \Gamma)$), ième groupe d'homologie, (resp. de cohomologie) de X, relativement à Γ; $H_*(X, \Gamma)$, (resp. $H^*(X, \Gamma)$), somme directe des H_i, (resp. H^i), (X sera toujours un polyèdre connexe fini ou éventuellement localement fini à homologie de type fini).

Dx est le degré de l'élément homogène x, $P_p(X, t)$ le polynôme (ou la série) de Poincaré de X pour la cohomologie à coefficients dans K_p:

$$P_p(X, t) = \Sigma_{i \geqq 0} \dim. H^i(X, K_p) t^i;$$

enfin, on dira que X a de la p-torsion $(p \neq 0)$ si Tors. ${}_pH^*(X, Z) \neq 0$; par convention, X est toujours sans 0-torsion.

0. 3. Algèbre spectrale des espaces fibrés. Le système formé d'un espace E fibré de base B et de fibre typique F est noté (E, B, F) ou (E, B, F, q), q étant la projection de E sur B. Dans son algèbre spectrale (E_r) on a donc, sous des hypothèses convenables, toujours vérifiées dans ce travail,

$$E_2^{s,t} \cong H^s(B, H^t(F, \Gamma)),^3 \qquad (\Gamma, \text{ anneau de coefficients}),$$

[3] Plus précisément E_2 est, comme on sait, l'algèbre de cohomologie de B à coefficients dans le système local formé par les algèbres de cohomologie des fibres; dans tous les cas considérés dans ce travail, ce système sera simple, presque toujours soit parce que B est simplement connexe, soit parce que le groupe structural de la fibration est connexe.

et E_∞ est l'algèbre graduée associée à $H^*(E, \Gamma)$ convenablement filtrée; s, t, $s+t$ sont respectivement le degré-base, le degré-fibre et le degré total et sont notés DB, DF, D.

${}^nE_r$ est la somme directe des $E_r^{s,t}$, $(s+t=n)$.

$\kappa_{r+1}{}^r$ est la projection canonique des d_r-cocycles de E_r sur E_{r+1}, et $\kappa_b{}^a = \kappa_b{}^{b-1} \circ \cdots \circ \kappa_{a+1}{}^a$, $(a<b)$, est la projection sur E_b des éléments de E_a vérifiant $d_r\kappa_r{}^a x = 0$ $(a \leqq r < b)$.

$T^s =$ idéal des éléments de filtration $\geqq s$ dans $H^*(E, \Gamma)$; on pose $J^{s,t} = J^s \cap H^{s+t}(E, \Gamma)$, par conséquent $E^{s,t} = J^{s,t}/J^{s+t,t-1}$.

On sait que $E_r^{s,0}$, (resp. $E_r^{0,t}$), s'identifie à un quotient, (resp. à un sous-module), de $H^s(B, \Gamma)$, (resp. de $H^t(F, \Gamma)$).

La transgression admet plusieurs définitions équivalentes ([3], § 5); ici, nous n'utiliserons que la définition "spectrale": $x \in H^t(F, \Gamma)$ est transgressif si et seulement si il fait partie de $E_{t+1}{}^{0,t}$, et la transgression est en dimension t, l'homomorphisme $E_{t+1}{}^{0,t} \to E_{t+1}{}^{t+1,0}$ défini par la différentielle d_{t+1} de E_{t+1}.

Nous utiliserons sans commentaire toutes les propriétés usuelles de l'algèbre spectrale (*voir* par exemple [3], § 4).

0.4. Espaces universels et espaces classifiants. Soit G un groupe de Lie compact. Un espace *universel,* (resp. classifiant), pour G sera noté E_G, (resp. B_G).

L'algèbre spectrale universelle pour G est l'algèbre spectrale de la fibration (E_G, B_G, G); on a donc $E_2 = H^*(B_G, H^*(G, \Gamma))$ et E_∞ est triviale, (i. e. $E_\infty{}^{0,0} \cong \Gamma$, $E_\infty{}^{s,t} = 0$ pour $s+t>0$).

$x \in H^t(G, \Gamma)$ est *universellement transgressif* s'il est transgressif dans E_G; il l'est alors dans une large classe d'espaces fibrés principaux de groupe structural G, ([3], § 18).

Remarque. En fait, on définit tout d'abord classiquement des espaces $E(n, G)$ et $B(n, G)$ universel ou classifiant pour G et *pour la dimension* n qui devrait en principe figurer plus haut. Cependant on peut en toute rigueur aisément éviter de mentionner ce n (qui quoi qu'il en soit peut être supposé arbitrairement grand) par un passage à la limite convenable; par exemple on peut définir des espaces universel ou classifiant pour G et pour tout n comme limite inductive d'une suite convenable d'espaces $E(n, G)$ ou $B(n, G)$, ([6], No. 1); l'algèbre spectrale universelle et celle des fibrations rappelées en 0.6 sont alors prises en cohomologie singulière ou d'Alexander-Spanier à supports fermés quelconques. On peut aussi directement passer à la limite sur les algèbres de cohomologie et les algèbres spectrales comme cela est fait dans

[3], § 18 et 22. L'algèbre spectrale universelle sera alors par définition une algèbre spectrale (E_r), $(r \geqq 2)$, qui pour tout n est isomorphe pour $D < n - \dim G$ avec l'algèbre spectrale d'une fibration $(E(n, G), (B(n, G), G)$; cette définition est licite car si $E'(n, G)$ est un deuxième espace universel pour G et pour n, de base $B'(n, G)$, le diagramme (18. 3) de [3], écrit pour E et E' montre non seulement que $H^*(B(n, G), \Gamma)$ et $H^*(B'(n, G), \Gamma)$ sont canoniquement isomorphes jusqu'à n, mais encore, vu [3], § 4d, que les algèbres spectrales de $(E(n, G), B(n, G), G)$ et de $(E'(n, G), B'(n, G), G)$ sont canoniquement isomorphes pour $D < n - \dim G$.

0. 5. L'homomorphisme $\rho^*(U, G)$. Si U est un sous-groupe fermé de G, E_G est aussi universel pour U, d'où une application $\rho(U, G) : E_G/U = B_U \to E_G/G = B_G$; elle définit un homomorphisme $\rho^*(U, G) : H^*(B_G, \Gamma) \to H^*(B_U, \Gamma)$ qui sera désigné par $\rho^*_p(U, G)$ lorsque $\Gamma = K_p$.

En fait, on peut faire correspondre une application $\rho(f) : B_U \to B_G$ à tout homomorphisme $f : U \to G$, ([6], No. 1), mais nous n'aurons pas besoin de cette généralisation.

0. 6. Quelques fibrations.

(a) Si (X, Y, U, q) est un espace fibré principal à groupe structural U de Lie compact, il existe une fibration dans laquelle l'espace a même homologie que Y, la fibre type est X et la base est B_U, ([3], Théorème 22. 1) ; par abus de langage, nous parlerons de la *fibration* (Y, B_U, X, r) ; bien entendu, l'homomorphisme i^* déduit de l'inclusion d'une fibre n'est autre que q^*, et r^* est l'homomorphisme caractéristique de la fibration donnée.

Dans le cas où U est discret, la suite spectrale de cette fibration est tout à fait analogue (et probablement isomorphe) à la suite spectrale des revêtements finis réguliers de H. Cartan-J. Leray.

(b) Si U est un sous-groupe fermé de G, la projection $\rho(U, G)$ donne lieu à une fibration $(B_U, B_G, G/U, \rho(U, G))$.

(c) Si U est *invariant* dans G, il existe une fibration dans laquelle l'espace a l'homologie de B_G, la base et la fibre type sont respectivement $B_{G/U}$ et B_U, ([3], Prop. 22. 2, ou [6], No. 1) ; ici encore on parlera de la fibration $(B_G, B_{G/U}, B_U)$. L'homomorphisme i^* de cette fibration est $\rho^*(U, G)$.

0. 7. Groupes de Lie simples compacts connexes. Les différentes structures de groupes de Lie simples compacts connexes se notent classiquement $\boldsymbol{A_l}$, $(l \geqq 1)$, $\boldsymbol{B_l}$, $(l \geqq 2)$, $\boldsymbol{C_l}$, $(l \geqq 3)$, $\boldsymbol{D_l}$, $(l \geqq 4)$, $\boldsymbol{G_2}$, $\boldsymbol{F_4}$, $\boldsymbol{E_6}$, $\boldsymbol{E_7}$, $\boldsymbol{E_8}$. Chacun

de ces symboles représente une classe de groupes ayant des algèbres de Lie isomorphes; les quatre premières, les structures classiques, ont chacune un représentant linéaire bien connu, soit:

pour $\boldsymbol{A_l}$: $\boldsymbol{SU}(l+1)$ = groupe unitaire unimodulaire de $l+1$ variables complexes.

pour $\boldsymbol{C_l}$: $\boldsymbol{Sp}(l)$ = groupe unitaire de l variables quaternioniennes.

pour $\boldsymbol{B_l}$, resp. $\boldsymbol{D_l}$: $\boldsymbol{SO}(2l+1)$, (resp. $\boldsymbol{SO}(2l)$), groupe orthogonal unimodulaire de $2l+1$, (resp. $2l$), variables réelles.

Les quotients de $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$, $\boldsymbol{SO}(2n)$, par leurs centres, qui sont cycliques à resp. n, 2, 2 éléments, sont notés $\boldsymbol{PU}(n)$, $\boldsymbol{PSp}(n)$, $\boldsymbol{PSO}(2n)$. Les groupes $\boldsymbol{SU}(n)$ et $\boldsymbol{Sp}(n)$ sont simplement connexes, tandis que $\boldsymbol{SO}(n)$, $(n \geqq 3)$, admet un revêtement simplement connexe à deux feuillets, le groupe $\boldsymbol{Spin}(n)$, dont le centre est cyclique d'ordre 2 pour n impair, d'ordre 4 pour $n = 2m$, m impair, et isomorphe à $Z_2 + Z_2$ quand n est un multiple de 4, (*voir* E. Cartan, Annali di Matematica, t. 4 (1927), pp. 209-256).

Les représentants simplement connexes des structures exceptionnelles $\boldsymbol{G_2}$, $\boldsymbol{F_4}$, $\boldsymbol{E_6}$, $\boldsymbol{E_7}$, $\boldsymbol{E_8}$ ont des centres cycliques d'ordres respectifs 1, 1, 3, 2, 1, (E. Cartan, loc. cit.); en particulier il n'y a, à un isomorphisme près, *qu'un* groupe de structure $\boldsymbol{G_2}$, (resp. $\boldsymbol{F_4}$, resp. $\boldsymbol{E_8}$), que nous désignerons par le même symbole.

I. Le produit de Pontrjagin.

1. Cohomologie d'un *H*-espace. Dans la définition suivante, H est une algèbre de dimension finie sur un corps K_p, graduée par des sous-espaces H^k, $(k \geqq 0)$, anticommutative (i. e. $a \cdot b = (-1)^{st} b \cdot a$, $a \in H^s$, $b \in H^t$), munie d'un élément neutre engendrant H^0.

La hauteur de $x \in H$ est l'entier h tel que $x^{h-1} \neq 0$, $x^h = 0$; elle vaut toujours deux lorsque Dx est impair et que $p \neq 2$.

DÉFINITION 1.1. *Soient* $J = (x_i)$, $(1 \leqq i \leqq m,\ Dx_i \leqq Dx_j$ *si* $i \leqq j)$, *un système de générateurs homogènes de* H, *et* s_i *la hauteur de* x_i.

(a) J est de type (M) s'il est minimal et si s_k est au plus égale à la hauteur de toute élément de la forme $x_k + P(x_1, \cdots, x_{k-1})$, où P est un polynôme $(1 \leqq k \leqq m)$.

(b) J *est q-semi-libre,* (q nul ou premier), si les monômes $x_1^{r_1} x_2^{r_2} \cdots x_m^{r_m}$, $(0 \leqq r_i < s_i)$, forment une base d'espace vectoriel de H,

et si de plus Dx_i est impair pour $q = 0$, s_i est une puissance de q pour $q = 2$ ou bien $q \neq 0$, Dx_i pair. Les monômes précédents seront les *monômes basiques* associés à J.

(c) J *est q-simple* si les monômes $x_{i_1}^{r_1} x_{i_2}^{r_2} \cdots x_{i_k}^{r_k}$, $(i_1 < \cdots < i_k$; $1 \leqq k \leqq m$, $0 \leqq r_i < q$ pour Dx_i pair et $q \neq 0$, $r_i = 0, 1$ sinon), forment avec 1 une base d'espace vectoriel de H; ces monômes seront les *monômes basiques associés* à J. Un élément $x \in H$ sera dit *décomposable relativement à* J s'il est somme de monômes basiques qui sont produits d'au moins deux x_i.

Au lieu de 2-simple ou 0-simple, nous dirons en général *simple*, conformément à la définition 6.4 de [3]. On peut bien entendu étendre ces définitions aux algèbres H de type fini, avec quelques légères modifications, (cf. [3], Définitions 6.1, 6.2). Certaines d'entre elles équivalent à la notion d'algèbre extérieure et ne sont introduites que pour simplifier certains énoncés.

Il est clair que si J est q-semi-libre, les éléments x_i forment avec leurs puissances non nulles d'exposants $q, q^2, q^3, \cdots$ un système q-simple de générateurs, le *système q-simple associé à* J.

Si un système q-simple possède a générateurs de degrés impairs $m_1, \cdots, m_a$ et b générateurs de degrés pairs $n_1, \cdots, n_b$, le polynôme de Poincaré de H est:

$$P_p(H,t) = \prod_{i=1}^{i=a} (1 + t^{m_i}) \cdot \prod_{j=1}^{j=b} (1 + t^{n_j} + t^{2n_j} + \cdots + t^{(q-1)n_j}), \quad (q \neq 0)$$

$$P_p(H,t) = \prod_{i=1}^{i=a} (1 + t^{m_i}) \cdot \prod_{j=1}^{j=b} (1 + t^{n_j}), \quad (q = 0), \tag{1.2}$$

et il est clair que tout système q-simple de générateurs possède a éléments de degrés $m_1, m_2, \cdots, m_a$ et b éléments de degrés $n_1, \cdots, n_b$, et enfin que

$$\dim H = 2^a \cdot q^b \qquad (q \neq 0)$$

$$\dim H = 2^{a+b} \qquad (q = 0). \tag{1.3}$$

On a vu dans [3], (Théorème 6.1 et No. 7):

1.4. *Soit X un H-espace*[4] *qui soit un polyèdre fini connexe. Alors tout système de type (M) de générateurs de $H^*(X, K_p)$ est p-semi-libre.*

En particulier, $H^*(X, K_p)$ possède *toujours* un système p-semi-libre de générateurs.

---

[4] C'est à dire un espace muni d'un produit tel que les translations à gauche et à droite induisent des automorphismes de la cohomologie.

1.5. *$H^*(X, K_p)$ est l'algèbre extérieure d'un sous-espace gradué par des degrés impairs lorsque l'une des conditions suivantes est remplie*;

(a). *$p \neq 2$ et $H^*(X, K_p)$ a un système simple de générateurs* ([3], Prop. 6. 1b).

(b). *$p \neq 2$ et $\dim H^*(X, K_p)$ est une puissance de deux.*

(c). *$p = 2$ et $P_2(X, t) = \prod_{i=1}^{i=m} (1 + t^{n_i})$, ($n_i$ impair, $1 \leqq i \leqq m$).*

(d). *X est sans p-torsion* ([3], Prop. 7. 2.).[5]

Vu (1. 3), (a) et (b) sont équivalents, il nous reste donc à établir (c). Soit $J = (y_i)$, $(1 \leqq i \leqq k)$, un système 2-semi-libre de générateurs de $H^*(X, K_2)$; le système simple associé J' est d'après l'hypothèse formé de m éléments de degrés n_i impairs ; il ne peut donc contenir d'élément de la forme $y_i \cdot y_i$ ce qui implique que $y_i \cdot y_i = 0$, $(1 \leqq i \leqq k)$; ainsi J' est égal à J, ne comprend que des éléments de degrés impairs et de carrés nuls, et engendre bien une algèbre extérieure.

1. 6. *Soit X un H-espace qui soit un polyèdre fini connexe. Alors si X est sans torsion, $H^*(X, Z)$ est l'algèbre extérieure d'un sous-groupe abélien libre gradué par des degrés impairs* ([3], Prop. 7. 3.).

Remarques.

1. 7. Les propriétés 1. 5 (a), (b), (c) sont en fait valables pour toute algèbre de Hopf, (au sens de [3], Définition 6. 2) sur un corps parfait puisque (1. 4) vaut sous cette hypothèse, ([3], Théorème 6. 1).

1. 8. En utilisant (1. 2) et (1. 3), on obtient aisément la généralisation suivante de (1. 5b) : Soit $p \neq 2$. Si $H^*(X, K_p)$ possède un système q-simple de générateurs, $(q \neq p)$, alors c'est l'algèbre extérieure d'un sous-espace gradué par des degrés impairs.

1. 9. Les définitions précédentes s'étendent aisément à des algèbres quelconques ; par exemple :

DÉFINITION. *Soit H une algèbre unitaire sur un anneau A. L'ensemble (x_i), $(x_i \in H, i \in I, I$ ensemble totalement ordonné$)$ est un système simple de générateurs de H si ce dernier est la somme directe faible des sous-modules*

---

[5] Signalons à ce propos une légère erreur de [3] : La réciproque partielle à la Prop. 7. 2, énoncée dans démonstration en remarque, p. 143, est valable sous l'hypothèse $H^1(X, K_p) = 0$, et non pas seulement Tors. $H^2(X, Z) = 0$.

monogènes engendrés par 1 *et par les monômes* $x_{i_1} \cdot \cdot \cdot x_{i_k}$ $(i_1 < \cdot \cdot \cdot < i_k;$ $k = 1, 2, \cdot \cdot \cdot)$.

2. Le produit de Pontrjagin. Soit X un H-espace. Le produit h: $X \times X \to X$ définit un homomorphisme h_* (compatible avec les degrés totaux), de $H_*(X, K_p) \otimes H_*(X, K_p)$ dans $H_*(X, K_p)$; l'élément $h_*(a \otimes b)$, *le produit de Pontrjagin de* a *et* b sera noté $a \vee b$. Ce produit ajoute les degrés, est distributif, est associatif si h l'est, possède un élément neutre engendrant H_0 si h a un élément neutre.

Nous noterons $\langle\ ,\ \rangle$ la forme bilinéaire canonique sur le produit d'un espace vectoriel et de son dual; on sait que H^* et H_* sont en dualité de même que $H^* \otimes H^*$ et $H_* \otimes H_*$ et que

$$\langle x \otimes y, u \otimes v \rangle = \langle x, u \rangle \langle y, v \rangle ; \tag{2.1}$$

d'autre part h^* et h_* sont transposés l'un de l'autre, ce qui signifie:

$$\langle h^*(x), a \otimes b \rangle = \langle x, a \vee b \rangle. \tag{2.2}$$

Un élément homogène $x \in H^*(X, K_p)$ est dit *primitif* si

$$h^*(x) = x \otimes 1 + 1 \otimes x, \tag{2.3}$$

ce qui équivaut visiblement à dire que x est orthogonal aux éléments décomposables de $H_*(X, K_p)$; cette formule permet aussi de définir les éléments primitifs de la cohomologie relativement à un anneau principal A, car $H^*(X, A) \otimes H^*(X, A)$ est toujours canoniquement contenue dans $H^*(X \times X, A)$.

DÉFINITION 2.4. *Soit* θ *l'automorphisme de* $H^*(X, K_p) \otimes H^*(X, K_p)$ *qui transforme* $x \otimes y$ *en* $(-1)^{st}\ y \otimes x$, $(Dx = s, Dy = t)$; *nous dirons que* h^* *est symétrique si* $\theta \circ h^* = h^*$.

PROPOSITION 2.5. *Le produit de Pontrjagin est anticommutatif si et seulement si* h^* *est symétrique.*

Démonstration immédiate: $u \vee v = (-1)^{st} v \vee u$ pour tout $u \in H_s(X, K_p)$, $v \in H_t(X, K_p)$ équivaut à

$$\langle x, u \vee v \rangle = \langle x, (-1)^{st} v \vee u \rangle \text{ pour tout } u, v \text{ et } x \in H^{s+t}(X, K_p)$$

c'est à dire à

$$\langle h^*(x), u \otimes v \rangle = \langle h^*(x), (-1)^{st} v \otimes u \rangle$$

ou encore, puisqu'évidemment:

$$\langle \theta \cdot h^*(x), u \otimes v \rangle = \langle h^*(x), (-1)^{st} v \otimes u \rangle,$$

à

$\langle \theta \circ h^*(x), u \otimes v \rangle = \langle h^*(x), u \otimes v \rangle$ pour u, v, x quelconques donc à $\theta \circ h^* = h^*$.

Remarques.

2. 6. La démonstration prouve plus précisément que $u \vee v = (-1)^{st} v \vee u$ pour tout $u \in H_s$, $v \in H_t$ si et seulement si h^* est symétrique sur les éléments de degré $s + t$.

2. 7. h^* est en particulier symétrique si $H^*(X, K_p)$ possède un système de générateurs primitifs.

PROPOSITION 2. 8. *Les deux conditions suivantes sont équivalentes.*

(a). $H^*(X, K_p)$, *(resp.* $H^*(X, Z)$ *est sans torsion et), a un système simple de générateurs primitifs.*

(b). $H_*(X, K_p)$, *(resp.* $H_*(X, Z)$*), est anticommutative et est l'algèbre extérieure d'un sous-espace, gradué par des degrés impairs si* $p \neq 2$, *(resp. d'un sous-groupe abélien libre gradué par des degrés impairs).*

(a) *entraîne* (b). Soit (x_i), $(1 \leqq i \leqq m)$, un système simple de primitifs; les éléments

$$x(i_1, i_2, \cdots, i_k) = x_{i_1} \cdot x_{i_2} \cdot \cdots \cdot x_{i_k}, \qquad (i_1 < \cdots < i_k; 1 \leqq k \leqq m),$$

forment avec 1 une base additive de $H^*(X, K_p)$; soit $(v(i_1 \cdots i_k))$ la base duale, $v(i_1, \cdots i_k)$ correspondant bien entendu à $x(i_1, \cdots, i_k)$; nous voulons montrer que $v(i) \vee v(i) = 0$ et que $v(i_1, \cdots, i_k) = v(i_1) \vee \cdots \vee v(i_k)$, ce qui établira (b). Par hypothèse, les x_i sont primitifs, donc

$$h^*(x_{i_1} \cdots x_{i_k}) = \Sigma_{a_1 < \ldots < a_j} \epsilon_{a_1 \ldots a_j} x_{a_1} \cdots x_{a_j} \otimes x_{a_{j+1}} \cdots x_{a_k} \tag{2.9}$$

$\{\alpha_1, \cdots, \alpha_j\}$ étant une partie ordonnée de $\{i_1, \cdots, i_k\}$, $\{\alpha_{j+1}, \cdots, \alpha_k\}$ son complément ordonné, et $\epsilon_{a_1 \ldots a_n}$ le nombre d'inversions de la suite $\{\alpha_1, \cdots, \alpha_j, \alpha_{j+1}, \cdots, \alpha_k\}$. On en déduit immédiatement que $v(i) \vee v(i)$ et, par récurrence sur k, que $v(i_1, \cdots, i_k) - v(i_1) \vee \cdots \vee v(i_k)$ sont orthogonaux à $H^*(X, K_p)$, d'où les égalités annoncées.

Même démonstration pour les coefficients entiers.

(b) *entraîne* (a). Soit (v_i) une base homogène d'un sous-espace dont $H_*(X, K_p)$ est l'algèbre extérieure; les éléments

$$v(i_1, \cdots, i_k) = v_{i_1} \vee \cdots \vee v_{i_k}, \qquad (i_1 < \cdots < i_k; 1 \leqq k \leqq m)$$

forment avec 1 une base additive de $H_*(X, K_p)$, et comme $v_i \vee v_i = 0$ pour tout i, les produits d'au moins deux v_i engendrent le sous-espace des éléments décomposables de $H_*(X, K_p)$. Soit $(x(i_1, \cdots, i_k))$ la base duale; les éléments $x_i = x(i)$ sont orthogonaux aux éléments décomposables de $H_*(X, K_p)$, donc primitifs, et vérifient de nouveau (2. 9); cela permet de prouver par récurrence sur k que $x(i_1, \cdots, i_k) - x(i_1) \cdots x(i_k)$ est orthogonal à $H_*(X, K_p)$, donc que $x(i_1, \cdots, i_k) = x(i_1) \cdots x(i_k)$; ainsi (x_i) est un système *simple* de générateurs primitifs.

Même démonstration pour les coefficients entiers.

Remarque. Nous dirons que les systèmes (x_i) et (v_i) de la démonstration précédente sont en dualité.

2. 10. *Conditions suffisantes pour l'existence d'un système simple de générateurs primitifs.*

(a) Si le produit est associatif, il suffit que $H^*(X, K_p)$, (resp. $H^*(X, Z)$) soit l'algèbre extérieure d'un sous-espace (resp. d'un sous-groupe abélien libre) gradué par des degrés impairs. L'existence d'un système de primitifs est alors assurée par le théorème de Samelson ([17], ou [11], pp. 133-134); en fait le théorème n'y est formulé qu'en caractéristique zéro, mais la démonstration de Leray vaut sans changement en caractéristique quelconque, et même pour les coefficients entiers.

(b) La condition (a) est en particulier remplie lorsque X est sans p-torsion, (resp. sans torsion), d'après 1. 5(d) et 1. 6; comme un polyèdre fini ne peut avoir de p-torsion que pour un nombre fini de p, cela montre que, sauf éventuellement pour un nombre fini de p, $H_*(X, K_p)$ est une algèbre extérieure anticommutative.

(c) Soit $X = G$ un groupe de Lie compact connexe. Il suffit que $H^*(X, K_p)$ (resp. $H^*(X, Z)$ soit sans torsion et) possède un système simple de générateurs universellement transgressifs, puisque ces éléments sont primitifs, [3], Proposition 20. 1). Cette condition est toujours vérifiée si (a) l'est, ([3], Théorème 19. 1) et en fait lui est équivalente en caractéristique différente de deux (resp. sur Z). En caractéristique deux cependant elle vaut dans d'autres cas (voir Nos. 8, 17, 22), qui ne semblent pas justiciables d'une démonstration algébrique analogue à celle du Théorème de Samelson, mais elle n'est pas toujours réalisée (voir Nos. 15, 16).

Nous nous sommes bornés à formuler 2. 8 dans un cas simple, le seul utile dans la suite. Indiquons cependant qu'elle admet la généralisation suivante:

PROPOSITION 2.11. *Les deux conditions suivantes sont équivalentes*:

(a) $H^*(X, K_p)$ *a un système p-simple de m générateurs primitifs* (x_i).

(b) $H_*(X, K_p)$ *est anticommutative et possède un système p-semi-libre de m générateurs* (v_i), $(Dv_i = Dx_i)$, *la hauteur de* v_i *étant égale à p si* Dv_i *est pair, à deux sinon.*

Démonstration tout à fait analogue à celle de (2.8), et que nous ne détaillerons pas. On s'appuyera sur le fait que si x est primitif de degré pair

$$\langle x^n, v^n \rangle = n!(\langle x, v \rangle)^n, \tag{2.12}$$

(démonstration par récurrence sur n immédiate) ; on utilisera (2.9) lorsque les x_i sont de degrés impairs, que l'on complétera par l'égalité suivante, où $y_1 \cdot \cdots \cdot y_l$ sont des primitifs de degrés pairs

$$h^*(y_1^{r_1} \cdot \cdots \cdot y_l^{r_l}) = \Sigma_{0 \leqq a_i \leqq r_i} \binom{r_1}{a_1} \cdot \cdots \cdot \binom{r_l}{a_l} y_1^{a_1} \cdot \cdots \cdot y_l^{a_l} \otimes y_1^{r_1 - a_1} \cdot \cdots \cdot y_l^{r_l - a_l}. \tag{2.13}$$

PROPOSITION 2.14. *Soit* $J = (x_i)$, $(1 \leqq i \leqq m)$, *un système p-simple, (resp. simple), de générateurs de* $H^*(X, K_p)$ *dans lequel* x_i *est de degré impair pour* $i \leqq k$, *de degré pair pour* $k+1 \leqq i \leqq l$, *primitif pour* $1 \leqq i \leqq l$, $(1 \leqq k \leqq l \leqq m)$.

Un élément $x \neq 0$, *décomposable relativement à J, combinaison linéaire de monômes basiques dans lesquels n'interviennent que* $x_1, \cdots, x_l$ *n'est pas primitif.*

Soient y_i, $(1 \leqq i \leqq n)$, les monômes basiques associés à J qui sont produits des x_i, $(1 \leqq j \leqq l)$. Les monômes $y_{ij} = y_i \otimes y_j$, $(1 \leqq i, j \leqq n)$ sont donc linéairement indépendants, et vu, (2.9) et (2.13), $h^*(x)$ en est une combinaison linéaire ; supposons que le monôme

$$x_{i_1} \cdot \cdots \cdot x_{i_s} x_{j_1}^{r_1} \cdot \cdots \cdot x_{j_t}^{r_t}$$
$$(i_1 < \cdots < i_s \leqq k < j_1 < \cdots < j_t \leqq l, 1 \leqq r_i < p, s + t \geqq 2),$$

ait le coefficient $c \neq 0$ dans x ; on déduit alors de (2.9) et (2.13) que $h^*(x)$ contient

$$x_{i_1} \cdot \cdots \cdot x_{i_{s-1}} x_{j_1}^{r_1} \cdot \cdots \cdot x_{j_t}^{r_t} \otimes x_{i_s}$$

avec le coefficient c si $s \geqq 1$ et qu'il contient

$$x_{j_1}^{r_1 - 1} x_{j_2}^{r_2} \cdot \cdots \cdot x_{j_t}^{r_t} \otimes x_{j_1}$$

avec le coefficient $r_1 \cdot c$ si $s = 0$, et $h^*(x)$ n'est pas primitif ; démonstration analogue dans le cas du système simple.

COROLLAIRE 2.15. *Soit $J = (x_i)$ un système p-simple, ou simple, de générateurs primitifs de $H^*(X, K_p)$. Alors tout primitif est une combinaison linéaire des x_i.*

Si $J = (x_i)$ est un système p-simple, ou simple, il n'y a pas de raison en général pour qu'une nouvelle base de l'espace sous-tendu par les x_i soit encore un système p-simple, ou simple; cependant:

PROPOSITION 2.16. *Soient $J = (x_i)$ un système p-simple, (resp. simple), de générateurs primitifs de $H^*(X, K_p)$, P l'espace sous-tendu par les x_i. Alors toute base (y_i) de P est un système p-simple, (resp. simple), de générateurs de $H^*(X, K_p)$.*

Si $p \neq 2$ et si J est simple, $H^*(X, K_p)$ s'identifie à l'algèbre extérieure de P, la démonstration est immédiate et bien connue. Il reste donc à examiner le cas du système p-simple pour $p \neq 0$.

Si $p \neq 2$ et Dy_i est impair, $y_i \cdot y_i = 0$; sinon, comme on calcule mod. p, l'élément y_i^p sera aussi primitif, donc combinaison linéaire des x_j ou des y_j vu (2.15); il en résulte immédiatement que tout monôme basique associé à J est combinaison linéaire de monômes de la même forme écrits à l'aide des y_i; ces derniers, dont le nombre est égal à celui des monômes basiques de J, sont donc forcément linéairement indépendants, et forment avec 1 une base de $H^*(X, K_p)$; ainsi (y_i) est un système p-simple de générateurs.

Remarques.

2.17. Si $X = G$ est un groupe de Lie compact connexe, tout élément universellement transgressif est primitif, les Propositions 2.14, 2.15, 2.16 restent donc valables si on y remplace primitif par universellement transgressif.

2.18. Bien entendu, les résultats (2.8) à (2.16) et leur démonstration subsistent sans changement pour $H^*(X, K_p)$ ou $H^*(X, Z)$ de type fini.

3. Opérateurs sur la cohomologie d'un espace fibré principal définis par l'homologie du groupe structural. Soient G un groupe de Lie compact connexe opérant sur deux espaces X, Y, et $f: X \to Y$ une application continue commutant à G. Leray, ([12], [13]), a indiqué que l'on pouvait faire opérer $H_*(G, K_0)$ sur $H^*(X, K_0)$ et $H^*(Y, K_0)$ et sur l'algèbre spectrale de f. Nous utiliserons ces opérations au No. 16, aussi voulons-nous en donner ici la définition et les principales propriétés en caractéristique quelconque.

3.1. Soient A et B deux espaces vectoriels sur K_p, $\hat{B}$ le dual de B; on vérifie aisément que si l'on pose $\sigma_u(a \otimes b) = \langle b, u \rangle a$, $(a \in A, b \in B, u \in \hat{B})$,

on fait correspondre à u un homomorphisme σ_u de $A \otimes B$ dans A. Si de plus h^* est un homomorphisme de A dans $A \otimes B$ on en déduit un endomorphisme $\vartheta_u = h^* \circ \sigma_u$ de A; on notera $x \cdot \vartheta_u$ l'image de x par ϑ_u.

3. 2. Soient E un espace sur lequel G opère, $\zeta : E \times G \to E$ l'application ainsi définie; on a

$$H^*(E \times G, K_p) = H^*(E, K_p) \otimes H^*(G, K_p)$$

et nous appliquons 3. 1 au cas $A = H^*(E, K_p)$, $B = H^*(G, K_p)$ et $h^* = \zeta^*$; autrement dit, si

$$\zeta^*(x) = \Sigma_i x_i \otimes y_i, \ (x, x_i \in H^*(E, K_p), y_i \in H^*(G, K_p)),$$

on pose

$$x \cdot \vartheta_u = \Sigma_i \langle y_i, u \rangle x_i;$$

Comme

$$H_*(E \times G, K_p) = H_*(E, K_p) \otimes H_*(G, K_p)$$

ζ_* définit un accouplement de $H_*(E, K_p)$, $H_*(G, K_p)$ à $H_*(E, K_p)$, que l'on peut aussi appeler le produit de Pontrjagin, par

$$x \vee v = \zeta_*(x \otimes v), \ (x \in H_*(E, K_p), v \in H_*(G, K_p)),$$

et il est clair que ϑ_u est le transposé du produit de Pontrjagin, c'est à dire que

$$\langle x \cdot \vartheta_u, v \rangle = \langle x, v \vee u \rangle \text{ pour tout } v \in H_*(E, K_p).$$

Ces opérateurs ont les propriétés suivantes:

(a) Soit $Du = s$; ϑ_u est un homomorphisme de $H^k(E, K_p)$ dans $H^{k-s}(E, K_p)$ pour tout k.

(b) $\vartheta_0 = 0$; ϑ_1 est l'identité; $\vartheta_{au+bv} = a\vartheta_u + b\vartheta_v$, $(a, b \in K_p)$.

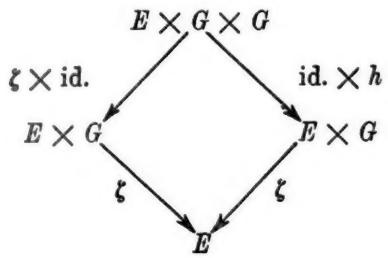
(c) $\vartheta_u \circ \vartheta_v = \vartheta_{u \cdot v}$.

(d) Soient X, Y deux espaces sur lesquels G opère, $f : X \to Y$ une application commutant à G. Alors $f^* \circ \vartheta_u = \vartheta_u \circ f^*$ pour tout $u \in H_*(G, K_p)$.

(e) Soit $u \in H_s(G, K_p)$ orthogonal aux éléments décomposables de $H^*(G, K_p)$ et soient x, y homogènes $\in H^*(E, K_p)$. Alors

$$(x \cdot y) \cdot \vartheta_u = x \cdot (y \cdot \vartheta_u) + (-1)^{s \cdot Dy}(x \cdot \vartheta_u) y.$$

(a) et (b) sont clairs; (c) résulte par un calcul facile de la commutativité du diagramme



qui exprime l'égalité $(e\cdot g)\cdot g' = e\cdot(g\cdot g')$; la propriété (d) se déduit de la commutativité du diagramme

$$\begin{array}{ccc} X\times G & \xrightarrow{\zeta_X} & X \\ f\downarrow & \downarrow \mathrm{id.} & \downarrow f \\ Y\times G & \xrightarrow{\zeta_Y} & Y\ . \end{array} \tag{3.21}$$

Enfin (e) est une conséquence immédiate de la définition, de l'égalité $f^*(x\cdot y) = f^*(x)\cdot f^*(y)$ et du fait que, vu l'existence d'un élément neutre dans G, on a, pour tout x homogène $\varepsilon\, H^*(E, K_p)$:

$$f^*(x) = x\otimes 1 + \Sigma x_i\otimes z_i,\ (x_i\,\varepsilon\, H^*(E,K_p),\ z_i\,\varepsilon\, H^*(G,K_p),\ 0<Dz_i<Dx).$$

3. 3. Soit maintenant (E, B, G, q) un espace fibré principal connexe, disons compact; G opère sur E, en laissant chaque fibre invariante, d'où un diagramme commutatif :

$$\begin{array}{ccc} E\times G & \xrightarrow{\zeta} & E \\ \downarrow q & & \downarrow q \\ B & \xrightarrow{\mathrm{id.}} & B \end{array} \tag{3.31}$$

ζ^* définit un homomorphisme de l'algèbre spectrale (E_r) de (E, B, G, q) dans celle de $(E\times G, B, G\times G, q)$, soit (E'_r) ; du reste ζ^* est biunivoque,

$$E'_r \cong \zeta^*(E_r)\otimes H^*(G,K_p) \cong E_r\otimes H^*(G,K_p), \qquad (r\geqq 2),$$

la différentielle d'_r de E'_r est égale à d_r sur $\zeta^*(E_r)$, nulle sur $1\otimes H^*(G,K_p)$, *voir* [3], lemme 20. 1, dont la démonstration montre de plus que

$$E'_r{}^{s,t} = \Sigma_{0\leqq i\leqq t} E_r{}^{s,t-i}\otimes H^i(G,K_p).$$

En prenant $A = E_r$, $B = H^*(G,K_p)$, $h^* = \zeta^*$, on définit selon 3. 1, un endomorphisme additif ϑ_u de E_r, pour tout $r\geqq 2$, qui a les propriétés suivantes :

(a). ϑ_u est un homomorphisme de $E_r{}^{j,k}$ dans $E_r{}^{j,k-s}$,

$(u\,\varepsilon\, H_s(G,K_p),\ r\geqq 2)$.

(b). $\vartheta_0 = 0$; $\vartheta_1 = \mathrm{id.}$; $\vartheta_{au+bv} = a\vartheta_u + b\vartheta_v$, $(a, b \in K_p)$.

(c). $\vartheta_u \circ \vartheta_v = \vartheta_{u \cdot v}$.

(d). $\vartheta_u \circ d_r = d_r \circ \vartheta_u$ et $\vartheta_u \circ \kappa_{r+1}{}^r = \kappa_{r+1}{}^r \circ \vartheta_u$.

(e). Sur E_2, on a $(b \otimes x) \cdot \vartheta_u = b \otimes (x \cdot \vartheta_u)$,
$(b \in H^*(B, K_p), x \in H^*(G, K_p))$.

(f). ϑ_u laisse invariants les idéaux J^s qui définissent la filtration de $H^*(E, K_p)$, et le diagramme suivant est commutatif.

$$\begin{array}{ccc} J^{j,k} & \xrightarrow{\vartheta_u} & J^{j,k-s} \\ \downarrow \pi & & \downarrow \pi \\ E_\infty{}^{j,k} & \xrightarrow{\vartheta_u} & E_\infty{}^{j,k-s} \end{array} \qquad (\pi \text{ projection canonique}).$$

La propriété (f) provient du fait que pour définir (E_r) et (E'_r) on filtre des couvertures fines bien choisies de E et $E \times G$ par le degré en les cochaînes de base, (*voir* par ex. [3], § 4, pour la définition précise), qui n'est pas altéré par le transposé de ζ ; la propriété (e) résulte de l'égalité $\zeta^*(b \otimes x) = b \otimes h^*(x)$ qui exprime que ζ est l'identité sur B et que sa restriction à une fibre $G \times \{g\}$ est le produit, (*voir* [3], démonstration du lemme 20. 1) ; les autres propriétés sont immédiates.

3. 4. En utilisant le diagramme (3. 21) au lieu de (3. 31) on pourrait aussi faire opérer $H_*(G, K_p)$ sur l'algèbre spectrale de f ; on obtiendrait ainsi les opérateurs du No. 3 de [13], sur lesquels nous n'insisterons pas. Le cas traité ici au No. 3. 3 correspond au No. 4 de [13].

3. 5. *Applications.* Nous considérerons ici la situation du diagramme (3. 21) dans le cas particulier $X = G$, l'application $\zeta_X = h$ étant bien entendu celle qui définit le produit de G.

THÉORÈME 3. 6. *Soient Y un espace sur lequel G opère, $f : G \to Y$ une application commutant à G, (opérant sur lui-même par les translations à droite). Si $H^*(G, K_p)$ possède un système simple de primitifs, alors $f^*(H^*(Y, K_p))$ est engendrée par des primitifs.*

En effet, vu 3. 2 (c), l'image de f^* est stable par rapport aux opérateurs ϑ_u de $H^*(G, K_p)$; il suffit donc de voir que toute sous-algèbre de $H^*(G, K_p)$ stable par rapport aux ϑ_u, (u parcourant $H_*(G, K_p)$), est engendrée par des primitifs, ce qui est élémentaire et laissé au lecteur.

THÉORÈME 3. 7. *On conserve les hypothèses de* 3. 6 *en supposant de plus*

que $p \neq 2$.[6] *Alors* $H^*(Y, K_p) = A \otimes \wedge P'$, *où* A *est une sous-algèbre annulée par* f^* *et où* $\wedge P'$ *est une sous-algèbre que* f^* *applique biunivoquement dans* $H^*(G, K_p)$.

Soient $x_1, \cdots, x_l$ une base de P dont les k premiers éléments engendrent l'image de f^*, ($Dx_i \leqq Dx_j$ si $i \leqq j \leqq k$) ; nous désignons par (u_i) la base duale et écrivons ϑ_i pour ϑ_{u_i} ; les opérateurs ϑ_i engendrent une algèbre extérieure, en particulier nous noterons $\wedge\vartheta'$ l'algèbre extérieure de $\vartheta_1, \cdots, \vartheta_k$. La démonstration de 3. 7 comprend trois parties.

(i) *A montrer*: Il existe $y_i \in H^*(Y, K_p)$ tel que $f^*(y_i) = x_i$ et que $y_i \cdot \vartheta_j = \delta_{ij}$, ($\delta_{ij}$ symbole de Kronecker, $1 \leqq i, j \leqq k$).

Vu que f^* est un isomorphisme en dimension zéro et que $y \cdot \vartheta_j = 0$ si $Dy < Dy_j$, on voit facilement que l'on peut trouver $\bar{y}_i \in f^{*-1}(x_i)$ tel que

$$\bar{y}_i \cdot \vartheta_j = \delta_{ij}, \qquad (1 \leqq i \leqq j \leqq k) ; \tag{3.71}$$

supposant avoir obtenu $y_1, \cdots, y_s$ vérifiant (i), nous voulons construire y_{s+1} ; l'élément

$$y_{s+1,s} = \bar{y}_{s+1} - (\bar{y}_{s+1} \cdot \vartheta_1 \cdots \cdots \vartheta_s) y_s \cdot y_{s-1} \cdots y_1$$

est annulé par le produit $\vartheta_1 \cdot \vartheta_2 \cdots \cdots \vartheta_s$, fait partie de $f^{*-1}(x_{s+1})$ car

$$f^*(\bar{y}_{s+1} \cdot \vartheta_1 \cdots \cdots \vartheta_s) = x_{s+1} \cdot \vartheta_1 \cdots \cdots \vartheta_s = 0$$

et vérifie

$$y_{s+1,s} \cdot \vartheta_{s+1} = 1, \qquad y_{s+1,s} \cdot \vartheta_j = 0 \qquad (s+1 < j \leqq k). \tag{3.72}$$

Admettons maintenant qu'il existe $y_{s+1,j} \in f^{*-1}(x_{s+1})$ vérifiant (3. 72), annulé par $\wedge^j(\vartheta_1, \cdots, \vartheta_s)$; si $j > 1$, alors

$$y_{s+1,j-1} = y_{s+1,j} - \Sigma_{i_1 < \ldots < i_j \leqq s} (y_{s+1,j} \cdot \vartheta_{i_1} \cdots \cdots \vartheta_{i_j}) y_{i_j} \cdots y_{i_1}$$

est dans $f^{*-1}(x_{s+1})$, vérifie (3. 72) et est annulé par $\wedge^{j-1}(\vartheta_1, \cdots, \vartheta_s)$. Finalement on arrivera à $y_{s+1,1}$ qui satisfait aux conditions imposées à y_{s+1} dans (i).

(ii). Les y_i sont de degrés impairs, de carrés nuls et leurs images par f^* engendrent $\wedge(x_1, \cdots x_k)$; par conséquent la sous-algèbre B engendrée par les y_i dans $H^*(Y, K_p)$ s'identifie à $\wedge(y_1, \cdots, y_k)$. Soit A l'ensemble des éléments de $H^*(Y, K_p)$ qui sont annulés par $\wedge^1\vartheta'$; c'est une sous-algèbre d'après 3. 2(e), qui fait visiblement partie du noyau de f^* ; nous voulons prouver que tout $y \in H^*(Y, K_p)$ est contenu dans la sous-algèbre engendrée par A et B, que nous notons $A \cdot B$.

---

[6] $H^*(X, K_p)$ est donc une algèbre extérieure à générateurs de degrés impairs d'après 1. 5(a).

Si y est annulé par $\wedge^1\vartheta'$, cela est vrai par définition de A; supposons-le établi pour tout élément annulé par $\wedge^s\vartheta'$ et soit y annulé par $\wedge^{s+1}\vartheta'$. Alors

$$y' = y - \Sigma_{i_1<\ldots<i_s\leqq k}\,(y\cdot\vartheta_{i_1}\cdot\cdots\vartheta_{i_s})y_{i_s}\cdot y_{i_{s-1}}\cdot\cdots y_{i_1}$$

est annulé $\wedge^s\vartheta'$, de même évidemment que les coefficients $(y\cdot\vartheta_{i_1}\cdot\cdots\cdot\vartheta_{i_s})$, et $y\,\varepsilon\,A\cdot B$ vu l'hypothèse d'induction d'où notre assertion, puisque tout élément est annulé par $\wedge^{k+1}\vartheta' = 0$.

(iii). Il reste encore à voir que $A\cdot B$ est isomorphe à $A\otimes B$; il est clair qu'il suffira pour cela d'établir:

(3.73). *Si l'élément*

$$z = \Sigma_{i_1<\ldots<i_s;\,1\leqq s\leqq k}\,a_{i_1,\ldots,i_s}y_{i_1}\cdot\cdots y_{i_s}$$

$(a_{i_1,\ldots,i_s}\,\varepsilon\,A)$, *est nul, alors chaque coefficient* $a_{i_1,\ldots,i_s}$ *est nul.*

Mais $z\cdot\vartheta_1\cdot\cdots\vartheta_k = a_{1,\ldots,k} = 0$ si $z = 0$; *cela étant*

$$z\cdot\vartheta_{i_1}\cdot\cdots\cdots\vartheta_{i_{k-1}} = a_{i_1,\ldots,i_{k-1}} = 0$$

et ainsi de suite, on démontre (3.73) par récurrence descendante sur s.

Remarques.

3.8. En caractéristique deux, la même démonstration montre que si $H^*(G, K_2)$ possède un système simple de générateurs primitifs, alors tout élément de $H^*(Y, K_2)$ peut se mettre sous la forme (3.73); la seule différence avec le cas 3.7 est que l'espace $\Delta(y_1,\cdots, y_s)$ n'est pas forcément une sous-algèbre.

3.9. Bien entendu, on a des résultats analogues en cohomologie entière lorsque $H^*(G, Z)$ et $H^*(Y, Z)$ sont sans torsion, (mêmes démonstrations).

3.10. Dans le cas particulier où f est l'inclusion de G dans un surgroupe de Lie compact connexe, 3.6 et 3.7 donnent respectivement les Prop. 21.2 et 21.3 de [3], elles-mêmes généralisations de résultats dûs à H. Samelson [17].

Si f est la projection de G sur un espace homogène G/U, le Théorème 3.7 généralise un résultat énoncé en caractéristique zéro par J. Leray [13].

II. Un théorème de transgression.

Le but essentiel de cette partie est la démonstration du Théorème 6. 1 qui peut s'envisager comme une réciproque au Théorème 19. 1 et à la Proposition 19. 1 de [3] ; tous ces résultats s'obtiennent par l'étude de l'algèbre spectrale universelle, mais tandis que ceux de [3] étaient cas particuliers de propositions purement algébriques, dans lesquelles il n'était même pas question d'espaces fibrés, nous utiliserons ici plus complètement les hypothèses topologiques, en faisant notamment intervenir les propriétés de l'algèbre de cohomologie d'un H-espace rappelées au No. 1.

Remarquons une fois pour toutes qu'en caractéristique zéro, l'hypothèse et la conclusion de 6. 1 sont toujours vraies, ([3], Théorème 19. 1), aussi supposerons-nous $p \neq 0$ dans les démonstrations des Nos. 4, 5, 6.

4. Enoncé d'une proposition et notations.

PROPOSITION 4. 1. *Soit G un groupe de Lie compact connexe et supposons que $H^*(B_G, K_p)$ est une algèbre de polynômes, pour un p donné. Alors $H^*(G, K_p)$ possède un système simple de générateurs universellement transgressifs, de degrés impairs si $p \neq 2$.*

Pour tout entier $j > 0$, nous prouverons, par récurrence sur j, l'existence d'un système M_j de générateurs de type (M) dans lequel les éléments de degré $\leqq j$ sont universellement transgressifs (et de degrés impairs pour $p \neq 2$), ce qui établira 4. 1.

Les éléments de $H^1(G, K_p)$ sont toujours tous universellement transgressifs ; on peut donc prendre pour M_1 n'importe quel système de type (M) ; supposant maintenant connaître un système M_k, $(k \geqq 1)$, nous nous proposons de construire M_{k+1} ; nous y parviendrons à la fin du No. 5, après avoir démontré une série de lemmes ; auparavant, introduisons quelques notations.

Notations.

(y_i) système de générateurs de $H^*(B_G, K_p)$.

D^j sous-espace des éléments décomposables de $H^j(B_G, K_p)$.

Q^j sous-espace sous-tendu par les y_i dont le degré vaut j ; il est donc isomorphe à $H^j(B_G, K_p)/D^j$.

$K_p[Q^j]$ algèbre de polynômes sur Q^j.

B_j sous-algèbre de $H^*(B_G, K_p)$ engendrée par 1 et par les y_i dont le degré est $\leqq j$; on a donc un isomorphisme naturel $B_j \cong K_p[Q^j] \otimes B_{j-1}$.

$N_k = (x_i)$: système p-simple associé à M_k.

S_i^j: Sous-espace de $H^*(G, K_p)$ engendré par 1 et par les monômes basiques associés à N_k qui sont produits d'éléments dont le degré vérifie $i \leqq Dx \leqq j$, S_i^j et $S_{j+1}{}^l$ sont donc isomorphes aux sous-espaces $S_i^j \otimes 1$ et $1 \otimes S_{j+1}{}^l$ de $S_i^j \otimes S_{j+1}{}^l$, qui est lui-même canoniquement isomorphe à S_i^l par un isomorphisme qui fait correspondre le produit $a \cdot b$ à $a \otimes b$, $(a \in S_i^j, b \in S_{j+1}{}^l)$.

(E_r) est l'algèbre spectrale sur K_p universelle pour G; on a donc $E_2{}^{s,t} = H^s(B_G, K_p) \otimes H^t(G, K_p)$ et E_∞ est triviale. On sait que dans toute algèbre spectrale $E_{j+1}{}^{j,0} \cong E_\infty{}^{j,0}$ et $E_{j+2}{}^{0,j} \cong E_\infty{}^{0,j}$, donc ici

$$E_{j+1}{}^{j,0} = E_{j+2}{}^{0,j} = 0 \qquad (j > 0). \tag{4.2}$$

Comme de plus $E_{j+1}{}^{j,0} \cong E_j^{j,0}/d_j(E_j^{0,j-1})$ et $E_{j+1}{}^{0,j-1} \cong d_j^{-1}(0) \cap E_j^{0,j-1}$ on voit que

(4.3) *d_j est un isomorphisme de $E_j^{0,j-1}$ sur $E_j^{j,0}$ $(j \geqq 2)$.*

5. Démonstration de la Proposition 4.1.

LEMME 5.1. *Soit $p \neq 2$. Alors $H^*(B_G, K_p)$ n'a que des éléments de degrés pairs et tout élément universellement transgressif $\neq 0$ est de degré impair.*

En caractéristique $\neq 2$, un élément de degré impair est de hauteur deux, $H^*(B_G, K_p)$, qui est par hypothèse une algèbre de polynômes, doit donc être engendrée par des éléments de degrés pairs. D'autre part le sous-espace $E_j^{0,j-1}$ des éléments universellement transgressifs de degré $j-1$ est isomorphe à $E_{j+1}{}^{j,0}$ d'après (4.3), donc nul si j est impair.

LEMME 5.2. *Soit $j \leqq k$. Supposons $A_j = K_p[Q^j] \otimes S_{j-1}{}^{j-1}$ muni d'une différentielle nulle sur $K_p[Q^j] \otimes 1$, qui applique l'espace des éléments de degré $j-1$ de $1 \otimes S_{j-1}{}^{j-1}$ isomorphiquement sur $Q^j \otimes 1$. Alors A_j a une cohomologie triviale.*

Remarquons tout d'abord qu'en caractéristique $p \neq 2$, il n'y a rien à démontrer si j est impair, et que pour j pair, $S_{j-1}{}^{j-1}$ est une algèbre extérieure à générateurs de degrés impairs.

A l'aide de la règle de Künneth, on se ramène immédiatement au cas où Q^j est de dimension 1, pour lequel la vérification est aisée; ce lemme résulte aussi pour $p \neq 2$ du lemme 11.1 et de la Proposition 11.1 de [3] et pour $p = 2$, des résultats correspondants dans le cas du système simple ([3], § 16).

LEMME 5.3. (a_r). *Pour $D \leqq k+3$, $\kappa_r{}^2$ est un isomorphisme de $B_r \otimes S_{r-1}{}^k$ dans E_r, $(2 \leqq r \leqq k+1)$.*

(b_r) $\kappa_r{}^2(B_r \otimes S_{r-1}{}^k)$ *contient* $E_r{}^{s,t}$ *pour* $s+t \leqq k+2, t \leqq k,$ $(2 \leqq r \leqq k+1)$.

(a_2) est évident, (b_2) résulte du fait que $E_2{}^{1,0} = 0$, (4. 2) ; supposons (a_j) et (b_j) démontrés pour un indice j, $(2 \leqq j \leqq k)$; il en résulte :

(5. 31) $\kappa_j{}^2(B_j \otimes S_{j-1}{}^k) \cong B_j \otimes S_{j-1}{}^k \cong K_p[Q^j] \otimes S_{j-1}{}^{j-1} \otimes S_j{}^k$, $(D \leqq k+3)$,

(5. 32) $E_j{}^{j,0} \cong Q^j$,

(5. 33) $E_j{}^{0,j-1} \subset \kappa_j{}^2(1 \otimes S_{j-1}{}^{j-1}) \cong S_{j-1}{}^{j-1}$ est le sous-espace sous-tendu par les x_i dont le degré vaut $j-1$.

(5. 34) $\kappa_j{}^2(B_j \otimes S_{j-1}{}^k)$ est stable pour d_j, et pour $D \leqq k+2$, d_j se transporte sur le dernier membre de (5. 31) en une différentielle bien déterminée par sa restriction à $E_j{}^{0,j-1}$, qui est un isomorphisme sur Q^j, vu (4. 3) et (5. 32), et par le fait qu'elle est nulle sur $K_p[Q^j] \otimes B_{j+1} \otimes S_j{}^k$.

Démonstration de (a_{j+1}) : d_j est nulle sur $B_{j+1} \otimes S_j{}^k$, donc pour que $\kappa_{j+1}{}^2 = \kappa_{j+1}{}^j \circ \kappa_j{}^2$ soit biunivoque sur ce sous-espace, $(D \leqq k+3)$, il suffit que $\kappa_{j+1}{}^j$ le soit, c'est à dire que

$$\kappa_j{}^2(B_{j+1} \otimes S_j{}^k) \cap d_j({}^sE_j) = 0 \text{ pour } s \leqq k+2 ;$$

soit $u \in {}^sE_j$; si $DFu \leqq k$, alors $u \in \kappa_j{}^2(B_j \otimes S_{j-1}{}^k) \cong B_j \otimes S_{j-1}{}^k$ et $d_j(u)$ est contenu dans l'idéal de Q^j, vu (5. 32) et (5. 34), et ne fait donc pas partie de $\kappa_j{}^2(B_{j+1} \otimes S_j{}^k)$; si $DFu > k$, alors $DBu = 0$ puisque $E_j{}^{1,0} = 0$ et que $Du \leqq k+2$, donc $DB(d_ju) = j$, et si $d_j(u) \neq 0$, il ne peut faire partie de $\kappa_j{}^2(B_{j+1} \otimes S_j{}^k)$, dans lequel les degrés base > 0 sont $\geqq j+1$.

Démonstration de (b_{j+1}). Un élément de $E_j{}^{s,t}$, $(s+t \leqq k+2, t \leqq k)$, peut s'écrire d'après (b_j) et (5. 31) :

(5. 35) $$h = u_1 \otimes v_1 + \cdots + u_n \otimes v_n + 1 \otimes v_{n+1},$$

($u_i \in K_p[Q^j] \otimes S_{j-1}{}^{j-1}$, $Du_i > 0$; $v_i \in B_{j+1} \otimes S_j{}^k$; $v_1, \cdots, v_n$ linéairement indépendants). L'espace $E_{j+1}{}^{s,t}$ est la projection, par $\kappa_{j+1}{}^j$ de l'espace des d_j-cocycles de $E_j{}^{s,t}$; pour prouver (b_{j+1}) il suffira donc de montrer que si $d_j(h) = 0$, chacun des n premiers termes de (5. 35) est un d_j-cobord ; en effet, on aura dans ce cas :

$$\kappa_{j+1}{}^j(h) = \kappa_{j+1}{}^j(1 \otimes v_{n+1}) \subset \kappa_{j+1}{}^j \circ \kappa_j{}^2(B_{j+1} \otimes S_j{}^k) = \kappa_{j+1}{}^2(B_{j+1} \otimes S_j{}^k).$$

Or si $d_j(h) = 0$, on doit avoir $d_j(u_i) = 0$, vu l'indépendance linéaire des v_i, $(1 \leqq i \leqq n)$; comme $Du_i > 0$, il existe $u'_i \in K_p[Q^j] \otimes S_{j-1}{}^{j-1}$ tel que $u_i = d_j(u'_i)$, (Lemme 5. 2 et (5. 34)), et alors

$$u_i \otimes v_i = d_j(u'_i \otimes v_i)$$

est bien un d_j-cobord, $(1 \leqq i \leqq n)$.

LEMME 5.4. *Soit x un élément de degré $k+1$ de M_k. Il existe $z \in S_1{}^k$, de degré $k+1$, tel que $x+z$ soit universellement transgressif.*

Supposons avoir trouvé $z_j \in S_1{}^k$ tel que $x+z_j$ soit un d_r-cocycle pour $r<j$; alors, vu (5.3):

$$h = d_j\kappa_j{}^2(x+z_j) \subset E_j{}^{j,k+2-j} \subset \kappa_j{}^2(B_j \otimes S_{j-1}{}^k) \cong B_j \otimes S_{j-1}{}^k$$

ou plus précisément, puisque $DB(h)=j$:

$$h = u_1 \otimes v_1 + \cdots + u_n \otimes v_n,$$

($u_i \in Q^j \otimes S_{j-1}{}^{j-1}$, $v_i \in S_j{}^k$, $(1 \leqq i \leqq n)$, $v_1, \cdots, v_n$ linéairement indépendants); évidemment $d_j(h)=0$, d'où $d_j(u_i)=0$ et (lemme 5.2), l'existence de $u'_i \in K_p[Q^j] \otimes S_{j-1}{}^{j-1}$ tel que $u_i = d_j u'_i$, $(1 \leqq i \leqq n)$; mais ici $DBu_i = j$ donc $DBu'_i = 0$, ce qui signifie que $u'_i \in S_{j-1}{}^{j-1}$; on en tire

$$d_j\kappa_j{}^2(x+z_j) = h = d_j(u'_1 \otimes v_1 + \cdots + u'_n \otimes v_n) = d_j(a),\ (a \in 1 \otimes S_{j-1}{}^k),$$

et ainsi $x+z_j-a = x+z_{j+1}$, $(z_{j+1} \in S_1{}^k)$, est un d_r-cocycle pour $r<j+1$, de manière analogue on montre l'existence de $z_2 \in 1 \otimes S_1{}^k$ tel que $d_2(x+z_2)=0$; on obtient donc par récurrence $z_{k+2} \in S_1{}^k$ tel que $x+z_{k+2}$ soit un d_r-cocycle pour $r<k+2$, autrement dit soit universellement transgressif.

LEMME 5.5. *Dans les notations du lemme 5.4, la hauteur de $x+z$ est égale à celle de x.*

Pour $p \neq 2$, x et $x+z$ sont de degrés impairs, (s'ils sont différents de zéro), d'après le lemme 5.1, donc tous deux de hauteur deux. Il reste à considérer le cas $p=2$.

Nous supposons les éléments du système simple $N_k = (x_i)$ associé à M_k numérotés de telle sorte que les l premiers soient les générateurs de M_k de degrés $\leqq k$ ainsi que leurs puissances non nulles d'exposants $2, 4, 8, \cdots$; la transgression commutant aux Sq^i, les éléments x_i, $(1 \leqq i \leqq l)$, sont tous universellement transgressifs. Soit $S \supset S_1{}^k$ le sous-espace engendré par 1 et par les monômes basiques

$$x_{i_1}x_{i_2}\cdots x_{i_a}, \qquad (i_1 < i_2 < \cdots < i_a \leqq l, 1 \leqq a \leqq l).$$

Comme l'ensemble des x_i, $(1 \leqq i \leqq l)$, contient le carré de chacun de ses éléments, lorsqu'il est $\neq 0$, on voit que S est une sous-algèbre dans laquelle, puisque nous calculons mod 2, le carré de tout élément décomposable relativement à $J=(x_i)$, $(1 \leqq i \leqq l)$, (Déf. 1.1c), est aussi décomposable relativement à J; en particulier, l'élément z de 5.4, de degré $k+1$, est forcément décomposable relativement à J, d'où vu (2.14) et (2.17):

(5. 51) Soit j un entier > 0; si $z^{2^j} \neq 0$, ce n'est pas un élément universellement transgressif.

La hauteur de x est une puissance de 2, (1. 4), disons 2^a; celle de $x + z$ est $\geqq 2^a$ par définition du système de type (M); il reste donc à montrer qu'elle est $\leqq 2^a$; or, comme nous calculons mod 2,

$$(x + z)^{2^a} = x^{2^a} + z^{2^a} = z^{2^a}$$

et z^{2^a} est universellement transgressif, en tant que puissance 2^a-ième d'un élément universellement transgressif, donc nul d'après (5. 51).

5. 6. *Construction du système M_{k+1}.* Soit $M_k = (h_i)$ le système donné; nous définissons un nouveau système (h'_i) par

$$h'_i = h_i \text{ si } Dh_i \neq k + 1$$

$$h'_i = h_i + z_i \text{ si } Dh_i = k + 1$$

z_i étant décomposable tel que h'_i soit universellement transgressif. Les z_i étant décomposables, (h'_i) est aussi minimal, et de plus tout polynôme en $h'_1, \cdots, h'_j$ est aussi un polynôme en $h_1, \cdots, h_j$, $(j = 1, 2, \cdots)$; un élément $h'_j + P(h'_1, \cdots, h'_{j-1})$ s'écrit donc aussi $h_j + Q(h_1, \cdots, h_{j-1})$ et sa hauteur est toujours $\geqq$ que la hauteur de h_j, donc que la hauteur de h'_j vu (5. 5); ainsi (h'_i) est un système de type (M) dans les éléments de degrés $\leqq k + 1$ sont universellement transgressifs; c'est le système M_{k+1} cherché.

6. Un théorème de transgression. En fait, tout l'essentiel est contenu dans la Proposition 4. 1, mais pour l'application que nous avons en vue, il sera plus commode d'utiliser l'énoncé suivant:

THÉORÈME 6. 1. *Soit G un groupe de Lie compact connexe et supposons que pour un p premier donné, $H^*(B_G, K_p)$ soit une algèbre de polynômes engendrée par m éléments $y_1, \cdots, y_m$, (forcément de degrés pairs si $p \neq 2$).*

Alors $H^*(G, K_p)$ possède un unique système simple de générateurs universellement transgressifs (u_i), $(i = 1, \cdots, m, Du_i = Dy_i - 1)$, tel que y_i soit une image de u_i par transgression dans E_G.

Prenons un système (h_i) de type (M) d'éléments universellement transgressifs, (Prop. 4. 1), et soit $N = (x_i)$ le système simple associé, (il faut bien remarquer que pour $p \neq 2$, les h_i sont de degrés impairs, et $N = M$ est bien un système simple, et non pas seulement p-simple). L'algèbre spectrale universelle vérifie relativement à N et pour D quelconque les propriétés que nous avons démontrées pour N_k et $D \leqq k + 2$, en particulier (5. 31), (5. 32),

(5.33). Ainsi l'espace $T(G, K_p)$ engendré par les éléments universellement transgressifs, et qui est la somme des $E_{j+1}^{0,j}$, $(j \geqq 1)$, admet les x_i comme base. Comme d_j est un isomorphisme de $E_j^{0,j-1}$ sur $Q^j = E_j^{j,0}$, ((4.3) et (5.32)), on voit que N compte m éléments et que si $y_a, \cdots, y_b$ est base de Q^j, l'espace $E_j^{0,j-1}$ possède une base $u_a, \cdots, u_b$ bien déterminée telle que

$$d_j \kappa_j^2 (1 \otimes u_i) = \kappa_j^2 (y_i \otimes 1), \qquad (a \leqq i \leqq b); \tag{6.2}$$

ce qui exprime justement que y_i est une image de u_i par transgression dans E_G. Le système (u_i) est complètement déterminé par (6.2), et comme c'est une base de $T(G, K_p)$ il constitue aussi un système simple de générateurs (*voir* (2.16) et (2.18)), d'où le théorème.

III. Groupes quotients des groupes classiques.

7. Sur l'algèbre spectrale des espaces fibrés. Dans ce No, A est un anneau principal, (E_r) l'algèbre spectrale sur A d'une fibration (E, B, F, q), où espace, base et fibres sont des polyèdres finis.

On notera i l'inclusion de F dans E, ι celle de $E_\infty^{s,0} = J^{s,0}$ dans $H^s(E, A)$, (cf. 0.3 pour les notations). Rappelons que

$$\iota \circ \kappa_\infty^2 = q^*; \qquad E_\infty^{0,t} \cong i^*(H^t(E, A)). \tag{7.1}$$

Nous nous proposons d'indiquer deux cas particuliers dans lesquels E_∞ est, (additivement et multiplicativement) isomorphe à $H^*(E, A)$; on supposera réalisée la condition suivante:

7.2. *E_∞ s'identifie au produit tensoriel (gauche sur A) de $U = \Sigma_{s \geqq 0} E_\infty^{s,0}$ et $V = \Sigma_{t \geqq 0} E_\infty^{0,t}$ par un isomorphisme qui applique le produit $x \cdot y$ sur $x \otimes y$, $(x \in E_\infty^{s,0}, y \in E_\infty^{0,t})$; en particulier $E_\infty^{s,t}$ s'identifie à $E_\infty^{s,0} \otimes E_\infty^{0,t}$.*

LEMME 7.3. *Sous l'hypothèse* (7.2), *J^s est l'idéal engendré par*

$$q^*(\Sigma_{i \geqq 0} H^{s+i}(B, A)).$$

Etant donné $h \in J^{s,t}$ il existe d'après (7.1) et (7.2) des éléments $x_i \in q^*(H^s(B, A)) = \iota(E_\infty^{s,0})$ et $y_i \in H^t(E, A)$ tels que

$$h - x_1 \cdot y_1 - \cdots - x_n \cdot y_n$$

soit dans $J^{s+1,t-1}$; pour $s + t$ fixé l'inclusion de $J^{s,t}$ dans l'idéal de

$$\iota(E_\infty^{s,0} + E_\infty^{s+1,0} + \cdots)$$

s'en déduit alors par récurrence descendante sur s; l'inclusion contraire résulte de $\iota(E_\infty^{s,0}) = J^{s,0} \subset J^s$ et de la règle $J^a \cdot J^b \subset J^{a+b}$.

PROPOSITION 7.4. *On suppose, outre* (7.2), *qu'il existe un homomorphisme* $\nu: V \to H^*(E, A)$ *tel que* $i^* \circ \nu$ *soit l'identité. Alors l'homomorphisme* $\iota \otimes \nu: U \otimes V \to H^*(E, A)$ *est un isomorphisme sur.*

C'est une légère généralisation d'une proposition de J.-P. Serre (*Annals of Math., 54* (1951), p. 473), qui s'établit exactement de la même façon, (rappelons que l'on filtre $U \otimes V$ par le degré en U et l'on montre que $\iota \otimes \nu$ induit un isomorphisme des algèbres graduées).

PROPOSITION 7.5. *On ajoute à* (7.2) *les hypothèses suivantes.*

(a) E *est un H-espace.*

(b) $A = K_p$ *est un corps de caractéristique* p.

(c) U, (resp. V), *possède un système p-semi-libre de générateurs* (u_i), $(1 \leqq l \leqq m, Du_i \leqq Du_j$ si $i \leqq j)$, (resp. (v_i), $(1 \leqq i \leqq n, Dv_i \leqq Dv_j$ si $i \leqq j)$ *et l'on a* $Du_m \leqq Dv_1$.

Alors $H^*(E, K_p)$ *est isomorphe à* $U \otimes V$.

Le système $(u_1, \cdots, v_n)$ est un système p-semi-libre de générateurs de $U \otimes V$, il est en particulier minimal et même de type (M). Soit $\hat{u}_i = \iota(u_i)$ et soit $\bar{v} \in H^*(E, K_p)$ tel que $i^*(\bar{v}_i) = v_i$; on voit aisément que c'est un système de générateurs de $H^*(E, K_p)$, ([3], Prop. 8.1); il est du reste minimal: En effet vu que $Du_m \leqq Dv_1$ la condition de minimalité est visiblement vérifiée par les $\hat{u}_i$, qui engendrent une sous-algèbre isomorphe à U; d'autre part, si $\bar{v}_i = P(\hat{u}_1, \cdots, \bar{v}_{i-1})$, alors

$$v_i = P(i^*(\hat{u}_1), \cdots, i^*(\bar{v}_{i-1})) = Q(v_1, \cdots, v_{i-1}), \qquad (\text{vu } i^*(\hat{u}_i) = 0),$$

ce qui est absurde. Soit enfin $\hat{v}_i = \bar{v}_i + P_i(\hat{u}_1, \cdots, \bar{v}_{i-1})$ de hauteur minimum parmi les éléments de la forme $\bar{v}_i + P(\hat{u}_1, \cdots, \bar{v}_{i-1})$, où P est un polynôme. Comme $\iota(U) \cong U$ et que $Du_m \leqq Dv_1$, les $\hat{v}_i$ forment avec les $\hat{u}_i$ un système de type (M) de générateurs de $H^*(E, K_p)$. Ce système est p-semi-libre vu l'hypothèse (a) et (1.4), et par conséquent

$$H^*(E, K_p) \cong \iota(U) \otimes \hat{V}, \qquad (\hat{V} \text{ sous-algèbre engendrée par les } \hat{v}_i).$$

Il nous faut encore montrer que $\hat{V} \cong V$. La hauteur de $\hat{u}_i$, soit r_i, est égale à celle de u_i; celle de $\hat{v}_i$, soit $\hat{s}_i$, est au moins égale à la hauteur de $i^*(\hat{v}_i) = v_i + P_i(i^*(u_1), \cdots, i^*(v_{i-1}))$, qui est $\geqq$ que la hauteur de v_i, soit s_i. Mais la définition de système p-semi-libre montre que

$$\dim H^*(E, K_p) = r_1 \cdots r_m \hat{s}_1 \cdots \hat{s}_n$$

$$\dim E_\infty = \dim U \otimes V = r_1 \cdots r_m s_1 \cdots s_n.$$

Comme ces deux dimensions sont égales, on doit avoir $s_i = \hat{s}_i$, $(1 \leqq i \leqq n)$, d'où l'isomorphie de $\hat{V}$ et V.

Remarques.

7. 6. Soit $v'_i = i^*(\hat{v}_i) = v_i + Q_i(v_1, \cdots, v_{i-1})$; sa hauteur est égale à celle de v_i, car elle est d'une part $\leqq$ que la hauteur de $\hat{v}_i$, qui vaut s_i, et d'autre part $\geqq$ que la hauteur de v_i, qui vaut s_i ; on en déduit sans difficulté que (v'_i) est aussi p-semi-libre, et que la correspondance $v'_i \to \hat{v}_i$ définit un homomorphisme $\nu : V \to H^*(E, K_p)$ tel que $i^* \circ \nu$ soit l'identité ; l'isomorphisme de 7. 5 peut donc être obtenu comme celui de 7. 4.

7. 7. Les résultats et démonstrations de ce No valent pratiquement sans changement lorsque E, B, F ont des cohomologies de type fini.

8. Homologie et cohomologie des groupes classiques.

Notations. Soient $x_1, \cdots, x_n$ n variables, σ_j la jème fonction symétrique élémentaire en les x_i, p un nombre premier.

On note $B_p^{k,j}(\sigma_1, \cdots, \sigma_j)$, $(j = i + k(p-1))$, le polynôme qui exprime la fonction symétrique de terme typique $x_1^p \cdots x_k^p x_{k+1} \cdots x_i$ en fonction des σ_j, et $b_p^{k,j}$ le coefficient de σ_j dans $B_p^{k,j}$, réduit mod. p. Par exemple on a ([6], 12. 3) :

$$b_p^{1,j} \equiv j \bmod. p. \tag{8. 1}$$

THÉORÈME 8. 2. *$\boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$ sont sans torsion. On a*

$$H^*(\boldsymbol{U}(n), Z) = H^*(\boldsymbol{S}_1 \times \boldsymbol{S}_3 \times \cdots \times \boldsymbol{S}_{2n-1}, Z)$$

$$H^*(\boldsymbol{SU}(n), Z) = H^*(\boldsymbol{S}_3 \times \boldsymbol{S}_5 \times \cdots \times \boldsymbol{S}_{2n-1}, Z)$$

$$H^*(\boldsymbol{Sp}(n), Z) = H^*(\boldsymbol{S}_3 \times \boldsymbol{S}_7 \times \cdots \times \boldsymbol{S}_{4n-1}, Z).$$

(Au point de vue additif, ce résultat est dû à Ehresmann [9] et Pontrjagin [16] ; pour une démonstration de 8. 2, voir [3], Proposition 9. 1) ; l'isomorphisme de 8. 2 est bien entendu valable additivement et pour le cup-produit. Par contre, au point de vue des puissances réduites on a [6], 11. 4 et 13. 5) :

THÉORÈME 8. 3. *Soit p un nombre premier. $H^*(\boldsymbol{U}(n), Z_p)$, (resp. $H^*(\boldsymbol{Sp}(n), Z_p)$), possède un système de n générateurs universellement transgressifs (h_i), $(Dh_i = 2i-1)$, (resp. (v_i), $Dv_i = 4i-1$), tels que*

$$\mathfrak{P}_p{}^k(h_i) = b_p{}^{k,j} \cdot h_j, \quad (1 \leqq i \leqq n, j = i + k(p-1)),$$

$$\mathfrak{P}_p{}^k(v_i) = 0 \text{ si } p = 2 \text{ et si } k \text{ est impair},$$

$$\mathfrak{P}_p{}^k(v_i) = (-1)^{k(p-1)/2} \cdot b_p{}^{k,2j} \cdot v_j \text{ sinon}, \quad (j = i + k(p-1)/2).$$

En combinant (2.8), (2.10), et (8.2), on voit que:

THÉORÈME 8.4. *Pour* $G = \boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$, *l'algèbre* $H_*(G, Z)$ *est une algèbre extérieure anticommutative, isomorphe à* $H^*(G, Z)$.

Avant de passer aux groupes orthogonaux, rappelons que la cohomologie de la variété $\boldsymbol{V}_{n,2}$ des vecteurs unité tangents à $\boldsymbol{S}_{n-1}$, déterminée par E. Stiefel (*Comm. Math. Helv.*, t. 8 (1935), pp. 3-50), est *pour n impair*:

(8.5) $H^0(\boldsymbol{V}_{n,2}, Z) = H^{2n-3}(\boldsymbol{V}_{n,2}, Z) = Z$; $H^{n-1}(\boldsymbol{V}_{n,2}, Z) = Z_2$; $H^i = 0$ *sinon.*

THÉORÈME 8.6. *Au point de vue additif,* $\boldsymbol{SO}(n)$ *a même cohomologie entière que le produit*

$$\boldsymbol{V}_{3,2} \times \boldsymbol{V}_{5,2} \cdots \times \boldsymbol{V}_{\bar{n},2} \qquad (\bar{n} \text{ plus grand entier impair} \leqq n),$$

multiplié encore par $\boldsymbol{S}_{n-1}$ *quand n est pair.*

(*voir* [8], [9], [16], [15], [3] Proposition 10.4). Les résultats suivants montrent que cet isomorphisme n'est pas valable au point de vue multiplicatif pour $n \geqq 5$.

THÉORÈME 8.7. $H^*(\boldsymbol{SO}(n), Z_2)$ *possède un unique système simple de générateurs universellement transgressifs* (h_i), $(1 \leqq i \leqq n-1, Dh_i = i)$, *liés par les relations*:

$$Sq^i h_j = \binom{j}{i} h_{i+j}, \quad (i \leqq j, i + j \leqq n-1); \quad Sq^i h_j = 0 \text{ sinon.}$$

([15], Theorems 5.3, 6.1, [3], Prop. 23.3, [4], Théorème 9.1, remarquons en passant que l'existence du système de générateurs universellement transgressifs pourrait aussi se déduire du fait que $H^*(B_{\boldsymbol{SO}(n)}, Z_2)$ est une algèbre de polynômes en les classes caractéristiques réduites $w_2, \cdots, w_n$, qui est établi dans [4], No. 8, sans recours à $H^*(\boldsymbol{SO}(n), Z_2)$, et du Théorème 6.1 du présent travail.)

Le théorème (8.7) montre en particulier que

8.8. $H^*(\boldsymbol{SO}(n), Z_2)$ *admet un système 2-semi-libre de générateurs* $h_1, h_3, h_5, \cdots$, *la hauteur* s_i *de* h_{2i-1} *étant la plus petite puissance de 2 telle que* $s_i(2i-1) \geqq n$.

En ce qui concerne la cohomologie mod p impair on a ([3], Prop. 10.2, [6], No. 14):

THÉORÈME 8.9. *Soit p un nombre premier impair. $H^*(\boldsymbol{SO}(2n+1), Z_p)$, (resp. $H^*(\boldsymbol{SO}(2n), Z_p)$), est une algèbre extérieure engendrée par n éléments $b_1, \cdots, b_n$, (resp. $b_1, \cdots, b_{n-1}, u_{2n}$), ($Db_i = 4i-1$, $Du_{2n} = 2n-1$), universellement transgressifs, vérifiant:*

$$\mathfrak{P}_p^k(u_{2n}) = 0 \text{ lorsque } k \geqq 1.$$

$$\mathfrak{P}_p^k(b_i) = (-1)^{k(p-1)/2} \cdot b_p^{k,2j} \cdot b_j, \quad (j = i + k(p-1)/2).$$

En utilisant (2.8), (2.10), (8.7) et (8.9), on obtient le théorème suivant, dû à Miller ([15], Theorems 4.2, 4.3):

THÉORÈME 8.10. *Soit $p \neq 2$. $H_*(\boldsymbol{SO}(n), Z_p)$ est une algèbre extérieure anticommutative, isomorphe à $H^*(\boldsymbol{SO}(n), Z_p)$.*

$H^*(\boldsymbol{SO}(n), Z_2)$ est une algèbre extérieure commutative, engendrée par $n-1$ éléments de degrés respectifs $1, 2, \cdots, n-1$.

Rappelons encore pour terminer ce No que certains des théorèmes précédents sont en fait des cas particuliers de théorèmes sur les variétés de Stiefel (réelles, complexes ou quaternioniennes). Pour ne pas trop allonger nous nous contenterons de mentionner un résultat qui sera utilisé plus loin ([15], Nos. 3, 5, 6, [3] Prop. 10.4 et [4] Théorème 9.1):

8.11. *Les coefficients de torsion de la variété de Stiefel $\boldsymbol{V}_{n,k}$ des k-repères orthonormaux de R^n sont tous d'ordre deux. La projection naturelle p_k: $\boldsymbol{SO}(n) \to \boldsymbol{V}_{n,k}$ induit un isomorphisme de $H^*(\boldsymbol{V}_{n,k}, Z_2)$ sur la sous-algèbre de $H^*(\boldsymbol{SO}(n), Z_2)$ qui est engendrée par les éléments h_i, ($n-k \leqq i \leqq n-1$; $1 \leqq k \leqq n-1$).*

9. Revêtements réguliers finis; cohomologie des groupes cycliques. Ce No est principalement consacré au rappel de notions et résultats pour la plupart bien connus.

Soit N un groupe fini. $H^*(B_N, \Gamma)$, (Γ anneau quelconque) s'identifie à l'algèbre de cohomologie de N relativement à Γ, au sens des groupes discrets. Supposons que N soit sous-groupe d'un groupe de Lie compact connexe G; il agit forcément trivialement sur sa cohomologie, par conséquent dans la fibration $(G/N, B_N, G)$ associée à $(G, G/N, N)$, (*voir* 0.5(a)), le système des $H^*(G, \Gamma)$ est simple et en particulier:

$$E_2 \cong H^*(B_N, K_p) \otimes H^*(G, K_p). \tag{9.1}$$

On sait que N a une cohomologie triviale relativement à K_p lorsque $p = 0$ ou lorsque p ne divise pas l'ordre de N; dans ce cas, E_2 est évidemment égale à E_∞, d'où la Proposition bien connue:

PROPOSITION 9.2. *Soient G un groupe de Lie compact connexe, N un sous-groupe discret de G, d'ordre n, et q la projection de G sur G/N.*

Alors $q^*: H^*(G/N, K_p) \to H^*(G, K_p)$ est un isomorphisme sur lorsque $p = 0$ ou lorsque p est premier à n.

COROLLAIRE 9.3. *Si p premier ne divise pas n, G et G/N sont simultanément avec ou sans p-torsion.*

Rappelons encore la cohomologie du groupe cyclique Z_n, (*voir* par exemple, S. Eilenberg, *Bull. Math. Soc.*, 55 (1949), pp. 3-37, No. 11):

PROPOSITION 9.4. *Z_n a une cohomologie triviale en caractéristique zéro, ou en caractéristique p si p ne divise pas n.*

Si p impair divise n, $H^*(B_{Z_n}, Z_p)$ est isomorphe au produit tensoriel gauche $\wedge(a) \otimes Z_p[b]$ $(Da = 1, Db = 2)$; on a le même résultat pour $p = 2$ lorsque $n \equiv 0 \bmod 4$.

Si $n = 2m$, m impair, $H^*(B_{Z_n}, Z_2) \cong Z_2[a]$, $(Da = 1)$.

PROPOSITION 9.5. *Soient $Z_m \subset Z_n$ des groupes cycliques d'ordres divisibles par p. Si $(m/n, p) = 1$, l'homomorphisme $\rho^*_p(Z_m, Z_n)$ est un isomorphisme sur, si m/n est divisible par p, l'homomorphisme $\rho^*_p(Z_m, Z_n)$ est un isomorphisme sur pour les éléments de degrés pairs et annule les éléments de degrés impairs.*

Soient $m = p^s q$, $n = p^t r$, $((q, p) = (r, p) = 1)$. On a donc

$$Z_m = Z_{p^s} \times Z_q, \qquad Z_n = Z_{p^t} \times Z_r,$$

$$B_{Z_m} = B_{Z_{p^s}} \times B_{Z_q}, \qquad B_{Z_n} = B_{Z_{p^t}} \times B_{Z_r}$$

et comme les deuxièmes facteurs des seconds membres ont une cohomologie triviale, l'homomorphisme $\rho^*_p(Z_m, Z_n)$ se ramène à l'homomorphisme ρ^*_p calculé pour les composantes p-primaires; il suffit donc d'examiner le cas $m = p^s$, $n = p^t$.

Notre assertion est évidente pour $s = t$; soit donc $s < t$ et posons $k = p^{t-s}$. Dans l'algèbre spectrale de $(B_{Z_n}, B_{Z_k}, B_{Z_m})$, (voir 0.5 (c)) on a

$$E_2 \cong H^*(B_{Z_k}, H^*(B_{Z_m}, Z_p))$$

et les coefficients sont ordinaires, car Z_k agit certainement trivialement sur $H^i(B_{Z_m}, Z_p) = Z_p$, donc

$$E_2 \cong H^*(B_{Z_k}, Z_p) \otimes H^*(B_{Z_m}, Z_p) = \wedge(a) \otimes Z_p[y] \otimes \wedge(b) \otimes Z_p[x]$$

$(Da = Db = 1, Dy = Dx = 2)$. Comme

$$\dim E_2^{1,0} = \dim E_\infty^{1,0} = 1 = \dim {}^1E_\infty$$

il faut que $E_3^{0,1} = 0$, ce qui n'est possible que si

$$d_2(1 \otimes b) = ky \otimes 1, \ (k \in Z_p, k \neq 0). \tag{9.51}$$

$E_2^{2,1}$ a l'élément $y \otimes b$ comme base, il ne contient donc pas de d_2-cocycle $\neq 0$, ni a fortiori de d_2-cobord $\neq 0$, par conséquent

$$d_2(1 \otimes x) = 0, \tag{9.52}$$

ces deux formules déterminent complèment d_2 et on calcule aisément (voir aussi Lemme 5.2) :

$$E_3 \cong \wedge(a) \otimes Z_p[x];$$

ainsi pour $r \geqq 3$, E_r ne contient pas de degré base $\geqq 2$, d_r est nulle et $E_3 \cong E_\infty$.

$Z_p[x]$ s'identifie à l'ensemble des éléments de degré fibre zéro de E_∞, c'est bien l'image de $\rho^*{}_p(Z_m, Z_n)$, (voir 0.5), d'où notre proposition dans le cas p impair. Pour $p = 2$, la démonstration est pratiquement identique à la précédente et laissée au lecteur.

10. Quotient d'un groupe de Lie par un sous-groupe cyclique.

Notation. Dans tout de No, Γ_m désigne un sous-groupe isomorphe à Z_m du groupe de Lie compact connexe G.

Nous voulons étudier la cohomologie mod. p du quotient G/Γ_m, (p/m), lorsque $H^*(G, Z_p)$ possède un système simple de générateurs universellement transgressifs.

LEMME 10.1. *Soient p un nombre premier, m un entier divisible par p. On suppose que $H^*(G, Z_p)$ est l'algèbre extérieure d'un sous-espace P gradué par des degrés impairs. Soit $2s$ le premier degré pour lequel $\rho^*{}_p(\Gamma_m, G)$ est non nul.*[7] *Alors dans l'algèbre spectrale sur Z_p de $(G/\Gamma_m, B_{\Gamma_m}, G, q)$:*

$$E_\infty \cong E_{2s+1} \cong H^*(B_{\Gamma_m}, Z_p)/(H^{2s}(B_{\Gamma_m}, Z_p)) \otimes \wedge P'$$

où, (pour un P convenable), P' est un sous-espace gradué de P tel que $\dim P' = \dim P - 1$, $\dim P'^{2s-1} = \dim P^{2s-1} - 1$. *Si de plus Γ_m est invariant dans G, l'entier s est une puissance de p.*

---

[7] La démonstration montrera l'existence de s.

On peut supposer que P est l'espace des éléments universellement transgressifs de $H^*(G, Z_p)$, ([3], Théorème 19. 1), par conséquent ([3], Prop. 22. 1) :

$$d_r\kappa_r{}^2(1 \otimes P^i) = 0 \text{ si } r \leqq i, \tag{10. 11}$$

$$d_i\kappa_i{}^2(1 \otimes P^{i-1}) = \kappa_i{}^2(\rho^*{}_p(\Gamma_m, G)\,(H^i(B_G, Z_p) \otimes 1)). \tag{10. 12}$$

Si $\rho^*{}_p(\Gamma_m, G)$ est nul sur les éléments de degrés $i < k$, ces formules montrent que $d_r\kappa_r{}^2(1 \otimes P) = 0$ pour $r < k$, donc que $d_r = 0$ et que $E_r \cong E_2$ pour $r < k$; mais E_2 est de dimension infinie, alors que E_∞ est de dimension finie, il existe donc un premier entier $2s$ tel que $\rho^*{}_p$ ne soit pas nul pour ce degré; l'espace $H^{2s}(B_{\Gamma_m}, Z_p)$ étant de dimension 1, le noyau P' de la restriction de d_{2s} à $\kappa_{2s}{}^2(1 \otimes P)$ est de dimension inférieure de 1 à celle de P; soit encore x un élément de P^{2s-1}, non contenu dans P'; on peut écrire

$$E_{2s} \cong E_2 \cong H^*(B_{\Gamma_m}, Z_p) \otimes \wedge(x) \otimes \wedge P',$$

$$d_{2s}\kappa_{2s}{}^2(1 \otimes P') = 0\,; \qquad d_{2s}\kappa_{2s}{}^2(1 \otimes x) = \kappa_{2s}{}^2(y \otimes 1) \neq 0$$

où y est un générateur de $H^{2s}(B_{\Gamma_m}, Z_p)$; on en tire aisément

$$E_{2s+1} \cong H^*(B_{\Gamma_m}, Z_p)/(H^{2s}(B_{\Gamma_m}, Z_p)) \otimes \wedge P',$$

et E_{2s+1}, qui ne contient plus d'élément non nul de degré base $\geqq 2s + 1$, vu (9. 4), est forcément égal à E_∞. Remarquons encore que s est la hauteur de l'élément $\sigma^*(y)$, où y est un générateur de $H^2(B_{\Gamma_m}, Z_p)$, et où σ^* est l'homomorphisme caractéristique de la fibration $(G, G/\Gamma_m, \Gamma_m, q)$.

Soit maintenant Γ_m invariant dans G. Si p est impair $\sigma^*(y)$ fait partie de tout système de type (M) de générateurs de $H^*(G/\Gamma_m, Z_p)$, et sa hauteur est une puissance de p, (1. 4). Si $p = 2$, $\sigma^*(y)$ ne fait pas forcément partie d'un système de type (M), mais on vérifie facilement à l'aide de (1. 4) que dans l'algèbre de cohomologie mod 2 d'un H-espace, les éléments de degré $\leqq 2$ ont toujours comme hauteur une puissance de deux.

COROLLAIRE 10. 2. *Le polynôme de Poincaré* mod p *de* G/Γ_m *est*

$$P_p(G/\Gamma_m, t) = P_p(G, t)\,(1 - t^{2s})\,(1 - t)^{-1} \cdot (1 - t^{2s-1})^{-1}.$$

PROPOSITION 10. 3. *On conserve les hypothèses du lemme* 10. 1 *et on suppose de plus que* p *est impair. Alors,*

$$H^*(G/\Gamma_m, Z_p) \cong \wedge(a) \otimes Z_p[y]/(y^s) \otimes \wedge P'.$$

En effet, d'après (9. 4) et le lemme, on a dans l'algèbre spectrale de $(G/\Gamma_m, B_{\Gamma_m}, G, q)$

$$E_\infty \cong E_{2s+1} \cong \wedge(a) \otimes Z_p[y]/(y^s) \otimes \wedge P',$$

et comme on est en caractéristique impaire, l'algèbre extérieure $\wedge P'$ se "remonte" dans $H^*(G/\Gamma_m, Z_p)$, et l'on peut appliquer (7.4).

PROPOSITION 10.4. *On suppose $p = 2$, Γ_m invariant dans G, et on conserve les hypothèses du lemme* 10.1. *Alors si $m \equiv 2 \bmod 4$:*

$$H^*(G/\Gamma_m, Z_p) \cong Z_2[a]/(a^{2s}) \otimes \wedge P', \qquad (Da = 1),$$

si $m \equiv 0 \bmod 4$ et si $H^1(G, Z_2) = 0$,

$$H^*(G/\Gamma_m, Z_2) \cong \wedge(a) \otimes Z_2[y]/(y^s) \otimes \wedge P', \qquad (Da = 1, Dy = 2).$$

D'après (9.4) et (10.1), on a dans l'algèbre spectrale de $(G/\Gamma_m, B_{\Gamma_m}, G, q)$

$$E_\infty \cong Z_2(a)/(a^{2s}) \otimes \wedge P' \text{ si } m \equiv 2 \bmod 4,$$

$$E_\infty \cong \wedge(a) \otimes Z_2[y]/(y^s) \otimes \wedge P' \text{ si } m \equiv 0 \bmod 4,$$

($Da = 1$, $Dy = 2$); comme nous supposons Γ_m invariant, on déduit l'isomorphie de $H^*(G/\Gamma_m, Z_p)$ avec E_∞ de (7.5), (qui s'applique aussi dans le deuxième cas puisque nous y supposons $H^1 = 0$).

PROPOSITION 10.5. *On suppose que $H^*(G, Z_2) = \triangle P$ possède un système simple de générateurs universellement transgressifs formant une base de P. Soit Γ_2 un sous-groupe d'ordre deux de G et soit s le premier degré pour lequel $\rho^*{}_2(\Gamma_2, G)$ est $\neq 0$.*[7] *Alors, dans l'algèbre spectrale de $(G/\Gamma_2, B_{\Gamma_2}, G)$, on a*

$$E_\infty \cong E_{s+1} \cong Z_2[a]/(a^s) \otimes \triangle P'$$

(P' sous-espace gradué de P, $\dim P' - 1$, $\dim P'^{s-1} = \dim P^{s-1} - 1$). Si Γ_2 est invariant, s est une puissance de deux.

Nous laissons au lecteur la démonstration, qui est pratiquement identique à celle de (10.1), si l'on tient compte du fait que toute base de P est un système *simple* de générateurs de $H^*(G, Z_2)$ d'après (2.16).

On pourrait aussi étudier le cas où m est divisible par 4, mais l'énoncé est plus compliqué car on doit envisager deux possibilités pour l'algèbre spectrale, suivant que s est pair ou impair, et nous ne le reproduirons pas ici. De même, si l'on cherche à étendre (10.4) au cas système simple de générateurs, on se heurte à quelques complications et nous nous bornerons au cas particulier suivant:

PROPOSITION 10.6. *On suppose que $H^*(G, Z_2)$ possède un système 2-semi-libre de générateurs universellement transgressifs $J = (x_i)$, $(1 \leqq i \leqq m$, $Dx_i < Dx_j$ si $i < j)$. Soit Γ_2 un sous-groupe invariant d'ordre deux de G et*

soit s le premier degré pour lequel $\rho^*_2(\Gamma_2, G)$ *est non nul. Alors J contient un élément de degré* $s-1$, *soit* x_j, *et* $H^*(G/\Gamma_2, Z_2) = Z_2[a]/(a^s) \otimes V$, $(Da = 1)$, *où V est isomorphe à la sous-algèbre de* $H^*(G, Z_2)$ *qui est engendrée par les* x_i *d'indices* $i \neq j$ *et par* x_j^2.

Soient J' le système simple associé à J et P le sous-espace sous-tendu par les éléments de J'; ces derniers sont universellement transgressifs et les formules (10. 11) et (10. 12) leur sont aussi applicables; par conséquent, tant que d_r est nulle sur J, sur J', (ce qui a lieu pour $r < s$), elle est identiquement nulle, d'où $E_s \cong E_2$. De plus d_s n'étant pas nulle il doit exister $x_j \in J$, non annulé par d_s, unique d'après l'hypothèse $Dx_a < Dx_b$ quand $a < b$. Le noyau P' de la restriction de d_s à P admet visiblement une base formée des éléments x_i, $(i \neq j)$, de x_j^2 et de leurs puissances non nulles d'exposants $2, 4, 8, \cdots$; comme J est 2-semi-libre, le sous-espace $\triangle P'$ est une sous-algèbre isomorphe à l'algèbre V décrite dans l'énoncé, et (10. 5) montre que dans l'algèbre spectrale de $(G/\Gamma_2, B_{Z_2}, G)$ on a

$$E_\infty \cong Z_2[a]/(a^s) \otimes V;$$

mais le système formé par les x_i, $(i \neq j)$ et x_j^2 est évidemment 2-semi-libre, et comme $Da = 1$, $H^*(G/\Gamma_2, Z_2)$ est isomorphe à E_∞ d'après (7. 5), ce qui établit notre proposition.

11. Groupes quotients des groupes classiques. Dans le No précédent, nous avons vu que l'étude de $H^*(G/\Gamma_m, Z_p)$ se ramène essentiellement à celle de $\rho^*_p(\Gamma_m, G)$ lorsque G possède un système simple de générateurs universellement transgressifs; dans le cas où G est classique, cette condition est toujours vérifiée et $\rho^*_p(\Gamma_m, G)$ se calcule aisément, comme nous allons le voir ici lorsque Γ_m est un sous-groupe du centre de G.

LEMME 11. 1. *Soient* Z_m *le sous-groupe des éléments d'ordre m d'un tore à une dimension* T^1, *p un nombre premier divisant m. Alors* $\rho^*_p(Z_m, T^1)$ *est biunivoque.*

En effet, T^1/Z_m est aussi un cercle, $H^*(B_{T^1}, Z_p)$ est une algèbre de polynômes à un générateur de degré deux, donc, vu (9. 4)

$$P_p(B_{Z_m}, t) = P_p(T^1/Z_m, t) \cdot P_p(B_{T^1}, t),$$

l'algèbre spectrale mod. p de $(B_{Z_m}, B_T, T^1/Z_m, \rho(Z_m, T^1))$ est triviale, et $\rho^*_p(Z_m, T^1)$ est donc biunivoque.

LEMME 11. 2. *Soient* $(Z_m)^n$ *le sous-groupe des éléments d'ordre m d'un*

tore T^n à n dimensions, et p un nombre premier divisant m. Alors $\rho^*_p((Z_m)^n, T^n)$ est biunivoque.

C'est une conséquence évidente de (11.1). Remarquons encore que $\rho^*_2((Z_m)^n, T^n)$ se décrit facilement: Si $p = 2$ et si $m \equiv 2 \bmod 4$,

$$H^*(B_{(Z_m)^n}, Z_2) \cong Z_2[x_1, \cdots, x_n], \qquad (Dx_i = 1),$$

et l'image de $\rho^*_p((Z_m)^n, T^n)$ est la sous-algèbre des polynômes en les x_i^2. Si $p \geqq 3$ ou bien si $p = 2$ et si $m \equiv 0 \bmod 4$:

$$H^*(B_{(Z_m)^n}, Z_p) \cong \wedge(b_1, \cdots, b_n) \otimes Z_p[y_1, \cdots, y_n], \quad (Db_i = 1, Dy_i = 2),$$

et l'on peut visiblement prendre comme deuxième facteur (qui n'est pas univoquement déterminé), l'image de $\rho^*_p((Z_m)^n, T^n)$.

Nous passons maintenant aux quotients des groupes classiques par des sous-groupes de leurs centres, (pour les notations, cf. 0.7).

THÉORÈME 11.3. *Soient n un entier > 0 et $s = 2^k$ la plus grande puissance de deux divisant n. Alors, (avec $Da = 1$, $Dx_i = i$):*

$$H^*(\boldsymbol{PSp}(n), Z_2) \cong Z_2[a]/(a^{4s}) \otimes \wedge(x_3, x_7, \cdots, \hat{x}_{4s-1}, \cdots, x_{4n-1}).^8$$

D'après (8.3), les hypothèses de (10.1) et (10.4) sont vérifiées, il nous suffit donc de prouver que

(11.31). *$4s = 2^{k+2}$ est le plus petit degré > 0 pour lequel $\rho^*_2(\Gamma_2, \boldsymbol{Sp}(n))$, est $\neq 0$, Γ_2 étant le centre de $\boldsymbol{Sp}(n)$.*

Soient $\boldsymbol{T}^n$ le tore maximal de $\boldsymbol{Sp}(n)$ que forment les matrices diagonales à coefficients complexes, $(\boldsymbol{Z}_2)^n$ le sous-groupe de ses éléments d'ordre 2; les inclusions:

$$\Gamma_2 \subset (\boldsymbol{Z}_2)^n \subset \boldsymbol{T}^n \subset \boldsymbol{Sp}(n),$$

montrent que

$$\rho^*_2(\Gamma_2, \boldsymbol{Sp}(n)) = \rho^*_2(\Gamma_2, (\boldsymbol{Z}_2)^n) \circ \rho^*_2((\boldsymbol{Z}_2)^n, \boldsymbol{T}^n) \circ \rho^*_2(\boldsymbol{T}^n, \boldsymbol{Sp}(n)).$$

L'homomorphisme $\rho^*_2(\boldsymbol{T}^n, \boldsymbol{Sp}(n))$ est biunivoque, et son image dans

$$H^*(B_{\boldsymbol{T}^n}, Z_2) \cong Z_2[u_1, \cdots, u_n], \qquad (Du_i = 2),$$

est l'ensemble des fonction symétriques en $u_1^2, \cdots, u_n^2$, ([3], § 29); par conséquent, vu (11.2), l'image de $\rho^*_2((\boldsymbol{Z}_2)^n, \boldsymbol{Sp}(n))$ dans

$$H^*(B_{(\boldsymbol{Z}_2)^n}, Z_2) \cong Z_2[x_1, \cdots, x_n], \qquad (Dx_i = 1),$$

est l'ensemble des fonctions symétriques en $x_1^4, x_2^4, \cdots, x_n^4$.

---

[8] Comme d'habitude, une variable surmontée d'un ^ doit être omise.

Il reste à calculer $\rho^*_2(\Gamma_2, (\mathbf{Z}_2)^n)$. Or, on peut identifier $\mathbf{Z}_2$ et $(\mathbf{Z}_2)^n$ aux premiers groupes d'homologie (entière ou mod. 2) de leurs espaces classifiants de manière à ce que l'inclusion $\Gamma_2 \subset (\mathbf{Z}_2)^n$ se transporte en la restriction de $\rho_*(\mathbf{Z}_2, (\mathbf{Z}_2)^n)$ à la dimension 1.[9] D'autre part $(\mathbf{Z}_2)^n$ est formé des matrices diagonales dont les coefficients valent ± 1, et l'élément non nul de Γ_2, est la matrice -1; si v, resp. $v_1, \cdots, v_n$ sont des systèmes de générateurs de $H_1(B_{\Gamma_2}, Z_2)$, resp. $H_1(B_{(\mathbf{Z}_2)^n}, Z_2)$, l'image de v par ρ_* sera $v_1 + \cdots + v_n$, (pour un choix convenable évident des v_i), d'où, x, resp. $x_1, \cdots, x_n$ étant les bases duales des premiers groupes de cohomologie

$$\rho^*_2(\Gamma_2, (\mathbf{Z}_2)^n)(x_i) = x, \qquad (1 \leqq i \leqq n),$$

$$\rho^*_2(\Gamma_2, (\mathbf{Z}_2)^n)(\sigma_i(x_1^4, \cdots, x_n^4)) = \binom{n}{i} x^{4i}$$

(σ_i ième fonction symétrique élémentaire) ; mais on sait que

$$\text{(11.32)} \qquad \binom{n}{i} \equiv 0 \bmod 2 \text{ pour } i < 2^k, \ \binom{n}{2^k} \not\equiv 0 \bmod 2, \qquad (n = 2^k \cdot n', n' \text{ impair}),$$

ce qui, joint aux propriétés déjà établies de $\rho^*_2((\mathbf{Z}_2)^n, \boldsymbol{Sp}(n))$, donne (11.31).

THÉORÈME 11.4. *Soient Γ_m le sous-groupe d'ordre m du centre Γ_n de $\boldsymbol{SU}(n)$, p un nombre premier divisant m et $s = p^k$ la plus grande puissance de p divisant n. Alors, avec $Dx = 1$, $Dy = 2$, $Dx_i = i$:*

$$H^*(\boldsymbol{SU}(n)/\Gamma_m, Z_p) \cong Z_p[y]/(y^s) \otimes \wedge(x_1, x_3, \cdots, \hat{x}_{2s-1}, \cdots, x_{2n-1})^{8}$$

pour $p \geqq 3$, ou bien $p = 2$, $m \equiv 0 \bmod 4$, et pour $p = 2$, $m \equiv 0 \bmod 4$:

$$H^*(\boldsymbol{SU}(n)/\Gamma_m, Z_2) \cong Z_2[x]/(x^{2s}) \otimes \wedge(x_3, x_5, \cdots, \hat{x}_{2s-1}, \cdots, x_{2n-1}).$$

D'après (8.3), (10.3), (10.4), il nous suffit de montrer que:

(11.41) *$2s = 2p^k$ est le premier degré > 0 pour lequel $\rho^*_p(\Gamma_m, \boldsymbol{SU}(n))$ est non nul.*

$\boldsymbol{SU}(n)$ est totalement non homologue à zéro dans $\boldsymbol{U}(n)$, relativement à des coefficients quelconques, donc $\rho^*_p(\boldsymbol{SU}(n), \boldsymbol{U}(n))$ est *sur*, ([3], Cor. à la Prop. 21.3), et puisque

$$\rho^*_p(\Gamma_m, \boldsymbol{U}(n)) = \rho^*_p(\Gamma_m, \boldsymbol{SU}(n)) \circ \rho^*_p(\boldsymbol{SU}(n), \boldsymbol{U}(n))$$

il suffit de faire voir que (11.41) vaut aussi pour $\rho^*_p(\Gamma_m, \boldsymbol{U}(n))$, ce qui va s'établir exactement comme (11.31). On part des inclusions

$$\Gamma_m \subset (\boldsymbol{Z}_m)^n \subset \boldsymbol{T}^n \subset \boldsymbol{U}(n),$$

[9] Pour une discussion plus détaillée d'un cas analogue, cf. [4], No. 3.

$\boldsymbol{T}^n$ étant le tore maximal des matrices diagonales de $\boldsymbol{U}(n)$, d'où l'on déduit

$$\rho^*_p(\Gamma_m, \boldsymbol{U}(n)) = \rho^*_p(\Gamma_m, (\boldsymbol{Z}_m)^n) \circ \rho^*_p((\boldsymbol{Z}_m)^n, \boldsymbol{T}^n) \circ \rho^*_p(\boldsymbol{T}^n, \boldsymbol{U}(n)).$$

$\rho^*_p(\boldsymbol{T}^n, \boldsymbol{U}(n))$ est biunivoque, et son image est l'ensemble des fonctions symétriques en $y_1, \cdots, y_n$, ([3], § 29). Si $p = 2$, $m \equiv 2 \bmod 4$, on démontre comme plus haut que l'image de $\sigma_i(u_1, \cdots, u_n)$ par $\rho^*_p(\Gamma_m, \boldsymbol{T}^n)$ est

$$\rho^*_2(\Gamma_m, \boldsymbol{T}^n)(\sigma_i(u_1, \cdots, u_n)) = \binom{n}{i} x^{2i}.$$

Si maintenant $p \geqq 3$, ou bien $p = 2$, $m \equiv 0 \bmod 4$

$$H^*(B_{(\boldsymbol{Z}_m)^n}, Z_p) \cong \wedge(a_1, \cdots, a_n) \otimes Z_p[y_1, \cdots, y_n], \quad (Da_i = 1, Dy_i = 2),$$

et l'image de $\rho^*_p((\boldsymbol{Z}_m)^n, \boldsymbol{U}(n))$ est, vu (11.2), l'ensemble des fonctions symétriques en les y_i. Enfin l'homomorphisme $\rho^*_p(\Gamma_m, (\boldsymbol{Z}_m)^n$ dans

$$H^*(B_{\Gamma_m}, Z_p) \cong \wedge(b) \otimes Z_p[y], \qquad (Db = 1, Dy = 2),$$

est de nouveau transposé de l'inclusion, applique les b_i sur b et les y_i sur y, donc

$$\rho^*_p(\Gamma_m, (\boldsymbol{Z}_m)^n)(\sigma_i(y_1, \cdots, y_n)) = \binom{n}{i} y^i,$$

et (11.41) se déduit des relations

$$\binom{n}{i} \equiv 0 \bmod p \text{ si } i < p^k, \ \binom{n}{p^k} \not\equiv 0 \bmod p, \quad (n = p^k \cdot n', (n', p) = 1).$$

THÉORÈME 11.5. *Soient n un entier > 0, $s = 2^k$ la plus grande puissance de deux divisant $2n$. Alors,*

$$H^*(\boldsymbol{PSO}(2n), Z_2) = Z_2[a]/(a^s) \otimes V, \qquad (Da = 1),$$

où V est une algèbre unitaire ayant un système simple de $2n-2$ générateurs x_i, $(1 \leqq i \leqq 2n-1$, $i \neq 2^k - 1$, $Dx_i = n)$, liés par les relations $x_i \cdot x_i = x_{2i}$ si $2i \leqq 2n-1$, $x_i \cdot x_i = 0$ sinon.

$\boldsymbol{PSO}(2n)$ est le quotient de $\boldsymbol{SO}(2n)$ par un sous-groupe Γ_2 à deux éléments; (8.8) montre les hypothèses de (10.6) sont réalisées et qu'il suffira de faire voir que:

(11.51). *s est le plus petit degré > 0 pour lequel $\rho^*_2(\Gamma_2, \boldsymbol{SO}(2n))$ est non nul.*

L'homomorphisme $\rho^*_2(\boldsymbol{SO}(2n), \boldsymbol{O}(2n))$, où $\boldsymbol{O}(2n)$ est le groupe orthogonal complet à $2n$ variables, est *sur*, ([4], Prop. 8.1), il suffit donc d'établir

(11.51) pour $\rho^*_2(\Gamma_2, \boldsymbol{O}(2n))$. Soit $\boldsymbol{Q}(2n)$ le sous-groupe des matrices diagonales de $\boldsymbol{O}(2n)$, il est isomorphe à $(Z_2)^{2n}$, donc

$$H^*(B_{\boldsymbol{Q}(2n)}, Z_2) \cong Z_2[x_1, \cdots, x_{2n}], \qquad (Dx_i = 1);$$

les inclusions $\Gamma_2 \subset \boldsymbol{Q}(2n) \subset \boldsymbol{O}(2n)$ montrent que

$$\rho^*_2(\Gamma_2, \boldsymbol{O}(2n)) = \rho^*_2(\Gamma_2, \boldsymbol{Q}(2n)) \circ \rho^*_2(\boldsymbol{Q}(2n), \boldsymbol{O}(2n)),$$

$\rho^*_2(\boldsymbol{Q}(2n), \boldsymbol{O}(2n))$ applique $H^*(B_{\boldsymbol{O}(2n)}, Z_2)$ sur l'ensemble des fonctions symétriques en les x_i, ([4], Théorème 5.1). Quant à $\rho^*_2(\Gamma_2, \boldsymbol{Q}(2n))$ c'est de nouveau en dimension un le transposé de l'inclusion, et comme dans le cas de $\boldsymbol{Sp}(n)$, on aura:

$$\rho^*_2(\Gamma_2, \boldsymbol{Q}(2n))(\sigma_i(x_1, \cdots, x_{2n})) = \binom{2n}{i} x^i$$

et (11.51) résulte des propriétés (11.32) des coefficients binomiaux.

Remarques.

11.6. Si n est impair, s est égal à deux, le théorème montre que l'on obtient $H^*(\boldsymbol{PSO}(2n), Z_2)$ à partir de $H^*(\boldsymbol{SO}(2n), Z_2)$ simplement en y remplaçant h_1 par un élément de carré nul; en particulier, $\boldsymbol{PSO}(2n)$ a aussi un système simple de $2n-1$ générateurs de degrés $1, 2, \cdots, 2n-1$. On pourrait du reste déduire $s = 2$ du fait que le groupe fondamental de $\boldsymbol{PSO}(2n)$, qui est isomorphe au centre de $\boldsymbol{Spin}(2n)$, est isomorphe à Z_4, (0.7).

Par contre si n est pair, s est > 2, (ce qui si l'on veut correspond au fait que $\pi_1(\boldsymbol{PSO}(2n)) = Z_2 + Z_2$), et $H^*(\boldsymbol{PSO}(2n), Z_2)$ a un système simple de $2n-2+k$ générateurs.

11.7. Nous n'avons étudié la cohomologie mod. p de G/Γ_m que dans le cas où p divise m, le seul qui présente quelque intérêt. Lorsque $(p, m) = 1$, il suffit d'appliquer (9.2).

IV. Le groupe des spineurs.

12. Cohomologie du groupe des spineurs (résultats).

THÉORÈME 12.1. (a). *Pour p impair, $\boldsymbol{Spin}(n)$ n'a pas de p-torsion et $H^*(\boldsymbol{Spin}(n), Z_p)$ est isomorphe à $H^*(\boldsymbol{SO}(n), Z_p)$.*

(b). $\boldsymbol{Spin}(n)$ *a de la 2-torsion si et seulement si $n \geqq 7$, et ses coefficients de torsion sont tous égaux à deux.*

(c). *Soit $s(n)$ l'entier tel que $2^{s(n)-1} < n \leqq 2^{s(n)}$. L'algèbre $H^*(\boldsymbol{Spin}(n),$*

Z_2) possède un système simple de $n - s(n)$ générateurs $u_1, \cdots, u_{n-s(n)-1}, u$; le dernier est de degré $2^{s(n)} - 1$ et la suite des degrés $Du_1, Du_2, \cdots, Du_{n-s(n)-1}$ s'obtient à partir de la suite $3, 4, \cdots, n-1$ en y supprimant toutes les puissances de deux. Ces éléments vérifient les relations suivantes:

$$Sq^i u_j = \binom{Du_j}{i} u_k \text{ si } i \leqq Du_j,\ i + Du_j = Du_k;\ Sq^i u_j = 0 \text{ sinon};\ u \cdot u = 0.$$

12.1(a) résulte de (8.6) et (9.3); donnons maintenant les valeurs des degrés Du_i, Du pour $n \leqq 9$.

n	$s(n)$	$Du_1, \cdots, Du_{n-s(n)-1}$	Du
3	2		3
4	2	3	3
5	3	3	7
6	3	3, 5	7
7	3	3, 5, 6	7
8	3	3, 5, 6, 7	7
9	4	3, 5, 6, 7	15

13. Cohomologie mod. 2. Nous diviserons la démonstration de 12.1(c) en trois parties: (i). Détermination d'un système simple de générateurs, (ii). Détermination des u_i vérifiant les Sq^i annoncés, (iii). Détermination d'un générateur u de carré nul complétant le système des u_i.

(i). *Un système simple de générateurs.* L'espace homogène $\boldsymbol{V}_{n,n-2} = \boldsymbol{SO}(n)/\boldsymbol{SO}(2)$ est aussi le quotient de $\boldsymbol{Spin}(n)$ par l'image réciproque de $\boldsymbol{SO}(2)$ dans $\boldsymbol{Spin}(n)$; cette dernière est connexe puisque $\boldsymbol{V}_{n,n-2}$ est simplement connexe, c'est un revêtement à deux feuillets de $\boldsymbol{SO}(2)$, donc un cercle, d'où une fibration $(\boldsymbol{Spin}(n), \boldsymbol{V}_{n,n-2}, \boldsymbol{S}_1, q)$ dont nous voulons étudier l'algèbre spectrale sur Z_2. On déduit immédiatement des propriétés de $H^*(\boldsymbol{V}_{n,n-2}, Z_2)$ rappelées au No. 8:

13.1. L'idéal engendré par h_2 possède, en tant qu'espace vectoriel sur Z_2, une base formée des monômes $h_{i_1} \cdot h_{i_2} \cdots \cdots h_{i_k}$, $(i_1 < i_2 < \cdots < i_k)$, où l'un des i_j au moins est une puissance de deux.

13.2. Les $n - s(n) - 1$ éléments h_i dont les degrés ne sont pas des puissances de 2 constituent un système simple de générateurs d'une sous-algèbre unitaire K de $H^*(\boldsymbol{V}_{n,n-2}, Z_2)$ qui, en tant qu'espace vectoriel, est un supplémentaire de l'idéal (h_2).

13. 3. La hauteur de h_2 est égale à $2^{s(n)-1}$; l'annulateur de h_2, soit Ann. h_2, est l'idéal engendré par l'élément:

$$h^* = h_2 \cdot h_4 \cdots \cdots h_{2^{s(n)-1}} = h_2^{2^{s(n)-1}-1}, \qquad (Dh^* = 2^{s(n)} - 2);$$

l'application $y \to y \cdot h^*$ est une application linéaire biunivoque de K sur Ann. h_2.

Nous pouvons maintenant passer à l'algèbre spectrale sur Z_2 de $(\boldsymbol{Spin}(n), \boldsymbol{V}_{n,n-2}, \boldsymbol{S}_1, q)$. On a tout d'abord:

$$E_2 \cong H^*(\boldsymbol{V}_{n,n-2}, Z_2) \otimes H^*(\boldsymbol{S}_1, Z_2),$$

$$E_3 \cong E_\infty.$$

Soit x l'élément non nul de $H^1(\boldsymbol{S}_1, Z_2)$; comme $\boldsymbol{Spin}(n)$ est simplement connexe, on doit avoir $E_3^{0,1} = 0$ et x ne peut être un d_2-cocycle, d'où

$$d_2(1 \otimes x) = h_2 \otimes 1,$$

ce qui détermine complètement d_2, comme on sait; les d_2-cocycles forment la sous-algèbre

$$C(E_2) = \text{Ann. } h_2 \otimes x + H^*(\boldsymbol{V}_{n,n-2}, Z_2) \otimes 1,$$

et les d_2-cobords la sous-algèbre: $d_2(E_2) = (h_2) \otimes 1$.

En tant qu'espace vectoriel, Ann. $h_2 \otimes x + K \otimes 1$ est un supplémentaire de $d_2(E_2)$ dans $C(E_2)$ d'après (13. 2); mais, vu (13. 3) et (13. 1), cette sous-algèbre admet un système simple de générateurs formé des $n - s(n) - 1$ éléments h_i dont le degré n'est pas une puissance de deux et d'un élément $h^* \otimes x$ de degré égal à $2^{s(n)} - 1$; comme elle est appliquée isomorphiquement sur E_3 par κ_3^2, on voit que $E_3 \cong E_\infty$, donc aussi ([3], Prop. 8. 1) que $H^*(\boldsymbol{Spin}(n), Z_2)$ possède un système simple de $n - s(n)$ générateurs dont les degrés sont égaux aux entiers compris entre 3 et $n - 1$ et non puissances de deux et à $2^{s(n)} - 1$, ce qui établit (i).

(ii). *Détermination des u_i.* Soient (u_i), $(1 \leqq i < n - s(n))$, les éléments $q^*(h_j)$, (j non puissance de 2), rangés par ordre de degrés croissants et soit u^* un représentant dans $H^*(\boldsymbol{V}_{n,n-2}, Z_2)$ de $h^* \otimes x$. Vu (8. 7) et (8. 11), $Sq^i u_j$ a bien la valeur énoncée dans 12. 1(c); de plus les u_i sont représentés dans E_∞ par les éléments $\kappa_\infty^2(h_i)$, (i non puissance de deux), qui forment avec $h^* \otimes x$ un système simple de générateurs de E_∞, par conséquent ([3], Prop. 8. 1), les u_i et u^* constituent un système simple de générateurs de $H^*(\boldsymbol{Spin}(n), Z_2)$.

(iii). *Le générateur* u. Posons $U = q^*(H^*(\boldsymbol{V}_{n,n-2}, Z_2))$ et $a(n) = 2^{s(n)} - 1$. De (i) et (ii) on tire que U est isomorphe à $H^*(\boldsymbol{V}_{n,n-2}, Z_2)/(h_2)$ et contient tous les éléments de degrés $< a(n)$ de $H^*(\boldsymbol{Spin}(n), Z_2)$, que $H^{a(n)}(\boldsymbol{Spin}(n), Z_2)/U^{a(n)}$ est de dimension 1, et que tout élément de degré $> a(n)$ de $H^*(\boldsymbol{Spin}(n), Z_2)$ est décomposable. Un système de type (M) de générateurs de $H^*(\boldsymbol{Spin}(n), Z_2)$ comprendra donc un système de type (M) de générateurs de U, soit J, complété par un élément de degré $a(n)$ soit u.

Nous voulons montrer que la hauteur k de u vaut 2. Comme (J, u) est 2-semi-libre d'après (1.4), on a:

$$(13.4)\quad P_2(\boldsymbol{Spin}(n), t) = P_2(U, t).\ (1 + t^{a(n)} + t^{2a(n)} + \cdots + t^{(k-1)a(n)});$$

mais nous savons déjà que les u_i, (resp. les u_i et u^*), forment un système simple de générateurs de U, (resp. de $H^*(\boldsymbol{Spin}(n), Z_2)$, d'où

$$(13.5)\qquad P_2(\boldsymbol{Spin}(n), t) = P_2(U, t).\ (1 + t^{a(n)}).$$

ce qui, joint à (13.4), donne bien $k = 2$.

u étant ainsi de carré nul, et (J, u) étant 2-semi-libre, on voit que $H^*(\boldsymbol{Spin}(n), Z_2)$ est isomorphe à $U \otimes \wedge(u)$, donc que (u_i, u) en est un système simple de générateurs.

14. Cohomologie entière. Nous savons déjà que $\boldsymbol{Spin}(n)$ n'a pas de p-torsion pour $p \neq 2$; d'autre part, 12.1(a), (c) et le No. 8 montrent que pour $n < 7$, $\boldsymbol{Spin}(n)$ a même polynôme de Poincaré mod. 2 qu'en caractéristique zéro, il n'a donc pas de 2-torsion; naturellement cela résulte aussi du No. 8 et des isomorphies classiques:

$$\boldsymbol{Spin}(3) \cong \boldsymbol{Sp}(1), \qquad \boldsymbol{Spin}(4) \cong \boldsymbol{Sp}(1) \times \boldsymbol{Sp}(1),$$
$$\boldsymbol{Spin}(5) \cong \boldsymbol{Sp}(2), \qquad \boldsymbol{Spin}(6) \cong \boldsymbol{SU}(4).$$

Par contre, pour $n \geqq 7$, $P_2(\boldsymbol{Spin}(n), t)$ contient t^5 et t^6, qui ne figurent pas dans $P_0(\boldsymbol{Spin}(n), t)$, et $\boldsymbol{Spin}(n)$ a bien de la 2-torsion. Pour terminer la démonstration de 12.1(b), (et de 12), il nous reste donc à prouver que les coefficients de torsion de Tors. $H^*(\boldsymbol{Spin}(n), Z)$ sont tous égaux à deux pour $n \geqq 7$.

14.1. Soit X un polyèdre fini. On sait que l'homomorphisme

$$Sq^1\colon H^k(X, Z_2) \to H^{k+1}(X, Z_2)$$

n'est autre que l'homomorphisme de Bockstein attaché à la suite exacte de coefficients:

$$0 \to Z \to Z \to Z/2Z \to 0$$

suivi de la réduction mod. 2; il en résulte que dim. $Sq^1(H^k(X, Z_2))$ est égale au nombre des coefficients de torsion de Tors. ${}_2H^{k+1}(X, Z)$ qui sont égaux à deux; d'autre part les formules

$$H^k(X, Z_2) \cong H^k(X, Z) \otimes Z_2 + \mathrm{Tor}(H^{k+1}(X, Z), Z_2)$$
$$H^k(X, Z_0) \cong H^k(X, Z) \otimes Z_0 \cong H^k(X, Z)/\mathrm{Tors}.\, H^k(X, Z) \otimes Z_0$$

montrent que dim. $H^*(X, Z_2)$ — dim. $H^*(X, Z_0)$ est égal au double du nombre des coefficients de torsion de Tors. ${}_2H^*(X, Z)$. Par conséquent:

(14. 11). *Pour que les coefficients de torsion de* Tors. ${}_2H^*(X, Z)$ *soient tous égaux à deux, il faut et il suffit que*

$$2 \cdot \dim. Sq^1(H^*(X, Z_2)) = \dim. H^*(X, Z_2) - \dim. H^*(X, Z_0)$$

ou encore, puisque le premier membre est toujours $\leqq$ *au second, que*:

$$2 \cdot \dim. Sq^1(H^*(X, Z_2)) \geqq \dim. H^*(X, Z_2) - \dim. H^*(X, Z_0).$$

D'après 12. 1(a), (c), et le No. 8, on a

$$\dim. H^*(\boldsymbol{Spin}(n), Z_2) = 2^{n-s(n)}$$
$$\dim. H^*(\boldsymbol{Spin}(n), Z_0) = 2^{b(n)} \qquad (b(n) = [n/2]),$$

il nous suffira donc d'établir que

(14. 12) $$2 \cdot \dim. Sq^1(H^*(\boldsymbol{Spin}(n), Z_2) \geqq 2^{n-s(n)} - 2^{b(n)}.$$

14. 2. Nous reprenons les notations du No. 13; on a donc

$$H^*(\boldsymbol{Spin}(n), Z_2) \cong U \otimes \wedge(u) \cong \Delta(u_1, \cdots, u_{n-s(n)-1}) \otimes \wedge(u),$$

où $U = q^*(H^*(\boldsymbol{V}_{n,n-2}, Z_2))$ est stable pour Sq^i; soit (x_i) une base d'un supplémentaire dans U du noyau de Sq^1; comme $H^1(\boldsymbol{Spin}(n), Z_2) = 0$, il est clair que $Sq^1(u) \in U$, donc que les vecteurs $Sq^1(x_i \otimes 1) = Sq^1(x_i) \otimes 1$ et les vecteurs

$$Sq^1(x_i \otimes u) = Sq^1(x_i) \otimes u + x_i \cdot Sq^1(u) \otimes 1$$

forment un système libre; par suite

$$\dim. Sq^1(H^*(\boldsymbol{Spin}(n), Z_2)) \geqq 2 \cdot \dim. Sq^1(U)$$

et (14. 12) résultera de

(14. 21) $$4 \cdot \dim. Sq^1(U) \geqq 2^{n-s(n)} - 2^{b(n)}.$$

14.3. Nous avons vu que

$$U = q^*(H^*(\boldsymbol{V}_{n,n-2}, Z_2) = H^*(\boldsymbol{V}_{n,n-2}, Z_2)/(h_2)\,;$$

on peut donc écrire $U = A \otimes B$ avec

$A = \otimes_i \Delta(h_{2^i-1})$, $\quad (2 \leqq i < s(n))$, pour n impair,

$A = \otimes_i \Delta(h_{2^i-1}) \otimes \Delta(h_{n-1})$, $\quad (2 \leqq i < s(n))$, pour n pair,

$B = \otimes_i \Delta(h_{m_i}, h_{m_i+1})$, $(5 \leqq m_i \leqq n-2$, $m_i + 1$ pair, non puissance de 2). Nous noterons $c(n)$ le nombre de facteurs $\Delta(h_{m_i}, h_{m_i+1})$ de B, donc

$$c(n) = b(n) - s(n) + 1 \qquad (n \text{ impair}),$$

$$c(n) = b(n) - s(n) \qquad (n \text{ pair}).$$

Nous avons rappelé au No. 8 les Sq^i de $\boldsymbol{V}_{n,n-2}$; en particulier

$$Sq^1(h_i) = h_{i+1} \text{ pour } i \text{ impair} \leqq n-2,\ Sq^1(h_i) = 0 \text{ sinon},$$

$$h_i \cdot h_i = h_{2i} \text{ si } 2i \leqq n-1,\ h_i \cdot h_i = 0 \text{ sinon},$$

ce qui montre que A et B sont des sous-algèbres stables par Sq^1, et en fait que $Sq^1(A) = 0$; par des calculs élémentaires, que nous ne reproduirons pas, on en déduit que (14.21) équivaut à

$$2 \cdot \dim. Sq^1(B) \geqq 2^{2c(n)} - 2^{c(n)}. \tag{14.31}$$

14.4. Soit B_k, (k impair $\geqq 5$), la sous-algèbre de B, stable pour Sq^1, qui est engendrée par les h_i d'indices $\geqq k$; elle possède un système simple de générateurs, formé d'un nombre pair, que nous noterons $2l(k)$, d'éléments h_i; en particulier

$$B_5 = B, \qquad l(5) = c(n)\,; \qquad l(k-2) = l(k) + 1\,;$$

l'inégalité (14.31) résultera donc de:

$$2 \cdot \dim. Sq^1(B_k) \geqq 2^{2l(k)} - 2^{l(k)}, \qquad (k = 5, 7, \cdots) \tag{14.41}$$

que nous démontrerons par récurrence sur $l(k)$, (donc par récurrence descendante sur k).

Cette inégalité est immédiate pour $l(k) = 1$, supposons-la vraie pour un indice $k > 5$, et soient, dans B_k, N_k le noyau de Sq^1, I_k son image, S_k un supplémentaire de N_k et P_k un supplémentaire de I_k dans N_k, (on a bien $I_k \subset N_k$ puisque $Sq^1 \circ Sq^1 = 0$). Soit encore s l'indice tel que $l(s) = l(k)+1$, autrement dit tel que

$$B_s = \Delta(h_s, h_{s+1}) \otimes B_k.$$

L'espace $Sq^1(B_s)$ contient:

$$V_1 = (1 + h_{s+1}) \otimes I_k \cong I_k + I_k$$

$$V_2 = \{x, x = h_s \cdot h_{s+1} \otimes Sq^1 u + 1 \otimes h_{s+1} \cdot h_{s+1} \cdot u, u \in S_k\} \cong S_k$$

$$V_3 = h_{s+1} \otimes P_k \cong P_k$$

$$V_4 = \{x, x = h_s \otimes Sq^1 v + h_{s+1} \otimes v, v \in S_k\} \cong S_k.$$

On vérifie facilement que des vecteurs $v_i \neq 0$, $(v_i \in V_i)$ sont linéairement indépendants, et comme d'autre part

$$\dim. S_k = \dim. I_k,$$

$$\dim. B_k = \dim. N_k + \dim. I_k = \dim. P_k + 2 \cdot \dim. I_k,$$

on voit que

$$\dim. Sq^1(B_s) \geqq 4 \cdot \dim. I_k + \dim. P_k = 2 \cdot \dim. I_k + \dim. B_k.$$

Mais $\dim. B_k = 2^{2l(k)}$ et par hypothèse d'induction,

$$2 \cdot \dim. I_k \geqq 2^{2l(k)} - 2^{l(k)},$$

d'où finalement

$$\dim. Sq^1 B_s \geqq 2^{2l(k)} - 2^{l(k)} + 2^{2l(k)}$$

$$2 \cdot \dim. Sq^1(B_s) \geqq 2^{2l(k)+2} - 2^{l(k)+1} = 2^{2l(s)} - 2^{l(s)}$$

c. q. f. d.

15. Éléments universellement transgressifs de $H^*(Spin(n), Z_2)$.

Proposition 15. 1. *Dans les notations de* 12. 1, *les éléments* (u_i), $(1 \leqq i < n - s(n))$, *sont universellement transgressifs.*

Soient α et β les projections canoniques:

$$Spin(n) \xrightarrow{\alpha} SO(n) \xrightarrow{\beta} V_{n,n-2};$$

mod. 2, β^* est biunivoque, applique h_i sur l'élément universellement transgressif de degré i de $H^*(SO(n), Z_2)$, (voir [3], § 10, 23), soit x_i; considérons d'autre part le diagramme

$$\begin{array}{ccc} E_{Spin(n)} & \xrightarrow{\lambda} & E_{Spin(n)}/\Gamma_2 \\ \downarrow \mu & & \downarrow \nu \\ B_{Spin(n)} & \xrightarrow{\text{id.}} & B_{Spin(n)} \end{array}$$

où Γ_2 désigne le centre de $Spin(n)$, et où λ, μ, ν sont les projections canoniques;

il est clair que ν fait de $E_{\boldsymbol{Spin}(n)}/\Gamma_2$ un espace fibré principal de groupe structural $\boldsymbol{SO}(n)$, dans lequel x_i est donc transgressif. Comme la restriction de λ à une fibre est α, l'image réciproque par λ^* d'une cochaine de transgression pour x_i dans la fibration de droite sera une cochaîne de transgression pour $\alpha^*(x_i) = \alpha^* \circ \beta^*(h_i) = q^*(h_i) = u_j$, ($j$ étant tel que $Du_j = i$), dans la fibration de gauche, ce qui montre justement que u_j est universellement transgressif.

Il reste à savoir si l'on peut compléter le système des u_i par un générateur universellement transgressif; nous verrons que cela n'est possible que si $n \leqq 9$; plus précisément:

PROPOSITION 15. 2. *$H^*(\boldsymbol{Spin}(n), Z_2)$ a un système simple de générateurs universellement transgressifs si et seulement si $n \leqq 9$.*

Compte tenu de 6. 1 et de la Prop. 19. 1 de [3], cela équivaut à:

PROPOSITION 15. 3. *$H^*(B_{\boldsymbol{Spin}(n)}, Z_2)$ est une algèbre de polynômes si et seulement si $n \leqq 9$.*

Pour étudier $H^*(B_{\boldsymbol{Spin}(n)}, Z_2)$, nous nous servirons de l'algèbre spectrale sur Z_2 de $(B_{\boldsymbol{Spin}(n)}, B_{\boldsymbol{SO}(n)}, B_{\Gamma_2})$ qui correspond à la fibration $(\boldsymbol{Spin}(n), \boldsymbol{SO}(n), \Gamma_2)$, (cf. 0. 6).

Remarquons que si E_∞ est une algèbre de polynômes, il en est évidemment de même pour $H^*(B_{\boldsymbol{Spin}(n)}, Z_2)$; d'autre part, si $H^*(B_{\boldsymbol{Spin}(n)}, Z_2)$ est une algèbre de polynômes, alors 6. 1 et 12. 1 montrent qu'elle possède $n - s(n)$ générateurs de degrés $Du_i + 1$, $(1 \leqq i < n - s(n))$, et $Du + 1$, autrement dit que sa série de Poincaré est:

$$(15.4) \qquad (1 - t^{2^{s(n)}})^{-1} \cdot \prod_{i=1}^{i=n-s(n)-1} (1 - t^{Du_i+1})^{-1};$$

la Prop. 15. 3 sera donc conséquence de la:

PROPOSITION 15. 5. *Soit (E_r) l'algèbre spectrale sur Z_2 de $(B_{\boldsymbol{Spin}(n)}, B_{\boldsymbol{SO}(n)}, B_{\Gamma_2})$. Alors pour $n \leqq 9$, E_∞ est une algèbre de polynômes et pour $n > 9$, E_∞ a pour le degré total une série de Poincaré différente de* (15. 4).

La démonstration de (15. 5) sera donnée au No. 15. 18, après une série de remarques préliminaires.

15. 6. Soit X un espace et soit $a = a_1 \cdot a_2 \cdot \cdots a_k$ un élément de degré j de $H^*(X, Z_2)$, $(a_i \in H^{m_i}(X, Z_2))$. Alors $Sq^{j-1}a$ est une somme de monômes qui ont la forme

$$a_1^2 \cdot \cdots a_{i-1}^2 (Sq^{m_i-1} a_i) \cdot a_{i+1}^2 \cdot \cdots a_k^2;$$

c'est une conséquence évidente des propriétés classiques des Sq^i.

15. 7. On sait que $H^*(B_{SO(n)}, Z_2) \cong Z_2[w_2, \cdots, w_n]$, $(Dw_i = i)$, est une algèbre de polynômes en les classes caractéristiques réduites w_i, dont les i-carrés sont donnés par les formules de Wu Wen Tsün:

$$Sq^i w_j = \Sigma_{0 \leqq t \leqq i} \binom{j-i+t-1}{t} w_{i-t} \cdot w_{j+t}, \qquad (i \leqq j, j = 2, \cdots, n),$$

avec les conventions: $w_1 = 0$; $w_k = 0$ si $k > n$; $\binom{a}{0} = 1$, $\binom{a}{b} = 0$ si $a < b$ et $b \neq 0$, (*voir* par ex. [4], No. 8). En particulier

$$(15.8) \qquad Sq^{j-1} w_j = w_{j-1} w_j + w_{j-2} w_{j+1} + \cdots + w_2 w_{2j-3} + w_{2j-1};$$

on en déduit que les éléments v_i définis par

$$v_i = w_i \quad \text{si } i \text{ n'est pas de la forme } 2^j + 1$$

$$(15.9) \qquad v_2 = w_2$$

$$v_{2^j+1} = Sq^{2^{j-1}}(v_{2^{j-1}+1}) = Sq^{2^{j-1}} \circ Sq^{2^{j-2}} \circ \cdots \circ Sq^2 \circ Sq^1(w_2)$$

$(0 < j < s(n))$, forment aussi un système algébriquement libre de générateurs de $H^*(B_{SO(n)}, Z_2)$.

Notations. $I(k)$ idéal de $H^*(B_{SO(n)}, Z_2)$ engendré par $v_2, \cdots, v_k$, ou, ce qui revient au même, par $w_2, \cdots, w_k$.

$L(k)$, idéal engendré par les éléments v_{2^j+1}, $(0 \leqq j < k)$.

$a(n) = 2^{s(n)} - 1$; $d(n) = 2^{s(n)-1}$.

15. 10. *Soit* $Q \in I(k)$, *(resp.* $Q \in L(k)$ *et* $k < s(n) - 1$*), de degré* j. *Alors* $Sq^{j-1}(Q) \in I_{2k-1}$, *(resp.* $Sq^{j-1}(Q) \in L(k+1)$*).*

C'est une conséquence évidente de (15. 6), (15. 8) et (15. 9).

15. 11. *On a* $v_{2^j+1} \equiv w_{2^j+1} \bmod I(2^{j-1})$, $(1 \leqq j < s(n))$.

La démonstration se fait aisément par récurrence sur j, à l'aide de (15. 8) et (15. 9), et est laissée au lecteur.

15. 12. *On a* $Sq^{d(n)} v_{d(n)+1} \in L(s(n))$ *pour* $n \leqq 9$.

La formule $Sq^2 v_3 = w_5 + w_2 w_3$ montre que (15. 12) est vérifiée pour $n < 5$ et que si $n \geqq 5$

$$v_5 \equiv w_5 \bmod L(2), \text{ ou encore } w_5 \in L(3);$$

par conséquent, d'après (15. 8) et (15. 10):

$$Sq^4 v_5 \equiv w_9 + w_2 w_7 + w_3 w_6 + w_4 w_5 \bmod L(3)$$

ce qui donne (15. 12) pour $5 \leqq n < 9$ et de plus, pour $n \geqq 9$;

$$v_9 \equiv w_9 \mod L(3) \text{ ou encore } w_9 \in L(4) ;$$

on en tire à l'aide de (15. 10) :

$$Sq^8 v_9 \equiv Sq^8 w_9 \mod L(4),$$

en particulier, si $n = 9$:

$$Sq^8 v_9 \equiv w_8 w_9 \mod L(4), \text{ donc } Sq^8 v_9 \in L(4).$$

Notons encore en passant que pour $n = 10$, on a :

$$\text{(15. 13)} \qquad \begin{aligned} Sq^8 v_9 &\equiv w_8 w_9 + w_7 w_{10} \mod L(4) \\ Sq^8 v_9 &\equiv v_8 v_9 + v_7 v_{10} \mod L(4). \end{aligned}$$

15. 14. Pour $n > 9$, $Sq^{d(n)} v_{d(n)+1} \notin L(s(n))$. D'après (15. 11), on peut écrire

$$v_{d(n)+1} = w_{d(n)+1} + Q + v_{d(n)/2} \cdot R$$

($Q \in I(2^{s(n)-2} - 1)$, R de degré $2^{s(n)-2} + 1$). On en déduit

$$\begin{aligned} Sq^{d(n)} v_{d(n)+1} = {} & Sq^{d(n)}(w_{d(n)+1} + Q) \\ & + v^2_{d(n)/2} \cdot Sq^{d(n)/2}(R) + R^2 \cdot Sq^{d(n)/2-1}(v_{d(n)/2}), \end{aligned}$$

d'où, vu (15. 8), (15. 10) et $n > 9$, donc $s(n) \geqq 4$:

$$\begin{aligned} Sq^{d(n)} v_{d(n)+1} &\equiv w_{d(n)} w_{d(n)+1} + w_{d(n)-1} w_{d(n)+2} && \mod I(d(n) - 2)) \\ Sq^{d(n)} v_{d(n)+1} &\equiv v_{d(n)} v_{d(n)+1} + v_{d(n)-1} v_{d(n)+2} && \mod I(d(n) - 2). \end{aligned}$$

$Sq^{d(n)} v_{d(n)+1}$ est un polynôme en les v_i, dans lequel le produit $v_{d(n)-1} \cdot v_{d(n)+2}$ figure ainsi avec le coefficient 1 ; il ne fait donc pas partie de l'idéal $L(s(n))$ pour $s(n) \geqq 4$, c. q. f. d.

15. 15. Nous pouvons maintenant étudier l'algèbre spectrale de la fibration $(B_{Spin(n)}, B_{SO(n)}, B_{\Gamma_2})$; puisque $B_{SO(n)}$ est simplement connexe, on a

$$E_2 \cong H^*(B_{SO(n)}, Z_2) \otimes Z_2[x] \qquad (Dx = 1),$$

ce qui peut s'écrire

$$E_2 \cong Z_2[v_2] \otimes Z_2[v_3, \cdots, v_n] \otimes \Delta(x) \otimes Z_2[x^2] ;$$

$B_{Spin(n)}$ étant simplement connexe, on a ${}^1E_\infty = 0$ et x n'est pas un d_2-cocycle, ce qui implique :

$$d_2(1 \otimes x) = v_2 \otimes 1 ;$$

évidemment

$$d_2(1 \otimes x^2) = 2 \cdot v_2 \otimes x = 0$$

et l'on voit, à l'aide de la règle de Künneth et de 5. 2, que:

$$E_3 \cong Z_2[v_3, \cdots, v_n] \otimes Z_2[x^2].$$

Notons encore que, la transgression commutant aux Sq^i, les éléments $x^2, x^4, x^8, \cdots$, sont transgressifs et de plus

$$\begin{aligned} &d_{2^{j+1}}\kappa_{2^{j+1}}{}^2(1 \otimes x^{2^j}) = \kappa_{2^{j+1}}{}^2(v_{2^{j+1}} \otimes 1), && (1 \leqq j < s(n)), \\ &d_{2d(n)+1}\kappa_{2d(n)+1}{}^2(1 \otimes x^{2d(n)}) = \kappa_{2d(n)+1}{}^2(Sq^{d(n)}v_{d(n)+1} \otimes 1). \end{aligned} \tag{15. 16}$$

15. 17. *A montrer: Pour* $2 \leqq r \leqq 2d(n) = 2^{s(n)}$, *on a* $E_r \cong E_{r+1}$ *si* r *n'est pas de la forme* $2^j + 1$, *et*

$$E_{r+1} \cong H^*(B_{SO(n)}, Z_2)/L(j+1) \otimes Z_2[x^{2(r-1)}] \text{ si } r = 2^j + 1, \quad (0 < j < s(n)),$$

Nous avons déjà calculé E_3; supposons (15. 17) établie pour E_k, où $k < 2d(n)$ alors

$$E_k \cong U \otimes V, \qquad U = \kappa_k{}^2(H^*(B_{SO(n)}, Z_2), \qquad V = \kappa_k{}^2(Z_2[x]),$$

et V est une algèbre de polynômes engendrée par x^s, où s est une puissance de deux. Si $k \neq 2^j + 1$, V est engendrée par un élément transgressif de degré $\neq k - 1$, donc $d_k(1 \otimes V) = 0$, d'où, ici, $d_k \equiv 0$ et $E_{k+1} \cong E_k$; si maintenant $k = 2^j + 1$, $(1 \leqq j < s(n))$, alors

$$E_k \cong E_{2^{j-1}+2} \cong H^*(B_{SO(n)}, Z_2)/L(j) \otimes Z_2[x^{2^j}]$$

$$E_k \cong H^*(B_{SO(n)}, Z_2)/L(j) \otimes \Delta(x^{2^j}) \otimes Z_2[x^{2^{j+1}}],$$

et, vu (15. 16), il en résulte de nouveau par (5. 2) et la règle de Künneth:

$$E_{k+1} \cong H^*(B_{SO(n)}, Z_2)/L(j+1) \otimes Z_2[x^{2^{j+1}}].$$

15. 18. *Démonstration de la Proposition* 15. 5. Nous savons par (15. 17) que

$$E_{2d(n)+1} \cong Z_2[v_2, \cdots, v_n]/L(s(n)) \otimes Z_2[x^{2d(n)}]. \tag{15. 19}$$

Soit $n \leqq 9$. Les formules (15. 12) et (15. 16) montrent que $d_{2d(n)+1}$ est nulle sur $Z_2[x^{2d(n)+1}]$, donc identiquement nulle sur $E_{2d(n)+1}$; comme $x^{2d(n)}$ est un d_r-cocycle pour tout $r > 2d(n) + 1$, les différentielles ultérieures seront aussi identiquement nulles, et $E_\infty \cong E_{2d(n)+1}$ est une algèbre de polynômes.

Si $n > 9$, $d_{2d(n)+1}$ n'est pas nulle sur

$$\kappa_{2d(n)+1}{}^2(1 \otimes x^{2d(n)})$$

vu (15. 14), (15. 16) et (15. 19); par conséquent la série de Poincaré de

$E_{2d(n)+2}$, et, *a fortiori*, celle de E_∞ seront majorées par celle de $E_{2d(n)+1}$, qui est justement égale à la série (15.4), ce qui termine la démonstration de la Prop. 15.5.

15.20. Nous ne chercherons pas à déterminer complètement la structure de $H^*(B_{Spin(n)}, Z_2)$, qui paraît être assez compliquée dans le cas général, et nous nous bornerons à décrire $H^*(B_{Spin(10)}, Z_2)$. D'après (15.13), (15.16) et (15.19), on a

$$E_{17} \cong Z_2[v_2, \cdots, v_{10}]/L(4) \otimes Z_2[x^{16}]$$

$$E_{17} \cong Z_2[v_4, v_6, v_7, v_8, v_{10}] \otimes \Delta(x^{16}) \otimes Z_2[x^{32}]$$

$$d_{17}\kappa_{17}{}^2(1 \otimes x^{16}) = \kappa_{17}{}^2(v_7 \cdot v_{10} \otimes 1)$$

d'où, par un calcul facile,

$$E_{18} = Z_2[v_4, v_6, v_7, v_8, v_{10}]/(v_7 \cdot v_{10}) \otimes Z_2[x^{32}],$$

(cela résulte aussi du reste du lemme 11.1 de [3] appliqué au cas du système simple, cf. [3], § 16). Par conséquent $E_{18} \cong E_{33}$ et

$$d_{33}\kappa_{33}{}^2(1 \otimes x^{32}) = \kappa_{33}{}^2 Sq^{16}(v_7 \cdot v_{10}) \otimes 1.$$

Par ailleurs

$$Sq^{16}(v_7 \cdot v_{10}) = Sq^{16}(w_7 \cdot w_{10}) = (Sq^6 w_7) \cdot w_{10}{}^2 + w_7{}^2 \cdot Sq^9 w_{10}$$

$$Sq^{16}(w_7 \cdot w_{10}) = (w_6 \cdot w_7 + w_5 \cdot w_8 + w_4 \cdot w_9 + w_3 \cdot w_{10}) \cdot w_{10}{}^2 + w_7{}^2 \cdot w_9 \cdot w_{10};$$

w_3, w_5, w_9 sont dans $L(4)$, (voir démonstration de 15.12), et les expressions de E_{17} et E_{18} écrites au début de (15.20) montrent que $L(4)$ et le produit $v_7 \cdot v_{10} = w_7 \cdot w_{10}$ font partie du noyau de $\kappa_{33}{}^2$; ainsi d_{33} est nulle sur l'image de x^{32} dans E_{33}, donc identiquement nulle; les différentielles ultérieures, toujours nulles sur un élément de degré 32, seront alors aussi identiquement nulles, et $E_\infty \cong E_{33} \cong E_{18}$. L'algèbre de polynômes $Z_2[x^{32}]$ peut évidemment se "remonter" de E_∞ à $H^*(B_{Spin(10)}, Z_2)$, on peut appliquer (7.4), ce qui donne finalement:

PROPOSITION 15.21. *L'algèbre $H^*(B_{Spin(10)}, Z_2)$ est isomorphe à*

$$Z_2[w_4, w_6, w_7, w_8, w_{10}, w_{32}]/(w_7 \cdot w_{10}), \qquad (Dw_i = i).$$

16. Sur l'homologie du groupe des spineurs. En combinant (2.8), (12.1) et (15.2), on obtient:

THÉORÈME 16.1. *Pour $p \neq 2$, l'algèbre $H_*(Spin(n), K_p)$ est une algèbre extérieure, anticommutative, isomorphe à $H^*(Spin(n), K_p)$.*

Si $n \leqq 9$, $H_*(\boldsymbol{Spin}(n), Z_2)$ *est une algèbre extérieure commutative à* $n - s(n)$ *générateurs* (v_i), $(1 \leqq i \leqq n - s(n))$; *le degré de* $v_{n-s(n)}$ *est égal à* $2^{s(n)} - 1$ *et la suite des degrés* $Dv_1, \cdots, Dv_{n-s(n)-1}$ *s'obtient en supprimant les puissances de deux dans la suite* $3, 4, \cdots, n-1$.

Nous nous proposons maintenant de déterminer le produit de Pontrjagin dans $H_*(\boldsymbol{Spin}(10), Z_2)$; à cet effet, nous aurons besoin de quelques propriétés de l'algèbre spectrale universelle sur Z_2 de $\boldsymbol{Spin}(10)$, qui font l'objet du Lemme 16. 2.

Notations. Il sera commode de mettre en indices les degrés des générateurs aussi désignerons-nous dorénavant le système simple de générateurs de $H^*(\boldsymbol{Spin}(10), Z_2)$ décrit dans le Théorème 12.1 par $u_3, u_5, u_6, u_7, u_9, u^*_{15}$, $(Du_i = i)$.

D est l'algèbre des éléments décomposables de $H^*(\boldsymbol{Spin}(10), Z_2)$ et $D^i = D \cap H^i(\boldsymbol{Spin}(10), Z_2)$.

LEMME 16. 2. *Il existe* $u_{15} \in H^{15}(\boldsymbol{Spin}(10), Z_2)$, *congru à* u^*_{15} mod. D [10] *tel que l'on ait, dans l'algèbre spectrale universelle sur* Z_2 *pour* $\boldsymbol{Spin}(10)$:

(a). u_{15} *est un* d_r*-cocycle pour* $2 \leqq r \leqq 9$ *et*

$$d_{10}\kappa_{10}^2(1 \otimes u_{15}) = \kappa_{10}^2(w_{10} \otimes u_8) \neq 0.$$

(b). $E_7 \cong Z_2[w_7, w_8, w_{10}, w_{32}]/(w_7 \cdot w_{10}) \otimes \Delta(u_6, u_7, u_9, u_{15})$.

(c). $\kappa_{10}^2(w_{10} \otimes 1) \neq 0$.

Nous utiliserons sans autre commentaire le fait que les u_i, $(i < 15)$, sont universellement transgressifs (15. 1) et que leurs images par transgression dans l'algèbre spectrale universelle sont $\neq 0$ (puisque E_∞ est triviale). Vu (15. 21), on a :

$$E_2 \cong Z_2[w_4, w_6, w_7, w_8, w_{10}, w_{32}]/(w_7, w_{10}) \otimes \Delta(u_3, u_5, u_6, u_7, u_9, u^*_{15}).$$

d'où évidemment, $d_2 = d_3 = 0$, $E_4 \cong E_2$ et

$$d_4\kappa_4^2(1 \otimes u_3) = \kappa_4^2(w_4 \otimes 1).$$

les quatre u_i suivants sont des d_4-cocycles, et, en ce qui concerne u^*_{15}, on peut en tout cas écrire :

$$h = d_4\kappa_4^2(1 \otimes u^*_{15}) = \kappa_4^2(a_1 \otimes b_1 + \cdots + a_s \otimes b_s),$$

---

[10] Comme le système (u_i) contient u_i^2 lorsqu'il est $\neq 0$, on vérifie aisément que tout élément congru à u mod. D forme avec les u_i un système simple de générateurs.

$(a_i \varepsilon w_4 \otimes \Delta(u_3)$, $a_i \neq 0$, $b_i \varepsilon \Delta(u_5, u_6, u_7, u_9)$; $b_1, \cdots, b_s$ linéairement indépendants) ; l'élément h est forcément un d_4-cocycle, d'où $d_4(a_i) = 0$ et $a_i = w_4 \otimes 1$, $(1 \leqq i \leqq s)$; ainsi, l'élément

$$u'_{15} = u^*_{15} + u_3 \cdot (b_1 + \cdots + b_s)$$

est un d_r-cocycle pour $2 \leqq r \leqq 4$, congru à u^*_{15} mod. D.[10]. On en tire par un calcul direct facile, ou à l'aide de (5. 2), que:

$$E_5 \cong Z_2[w_6, w_7, w_8, w_{10}, w_{32}]/(w_7 \cdot w_{10}) \otimes \Delta(u_5, u_6, u_7, u_9, u'_{15}) ;$$

par conséquent $d_5 = 0$, $E_5 \cong E_6$ et

$$d_6 \kappa_6^2 (1 \otimes u_5) = \kappa_6^2 (w_6 \otimes 1). \tag{16. 21}$$

Par un raisonnement en tous points semblable à celui qui nous a conduit à u'_{15} et à E_5, on en déduit l'existence de $x \varepsilon \Delta(u_6, u_7, u_9)$, de degré 15, tel que $u''_{15} = u'_{15} + x$ soit un d_5-cocycle pour $2 \leqq r \leqq 6$, et

$$E_7 \cong Z_2[w_7, w_8, w_{10}, w_{32}]/(w_7 \cdot w_{10}) \otimes \Delta(u_6, u_7, u_9, u''_{15}) ; \tag{16. 22}$$

il en résulte que

$$d_7 \kappa_7^2 (1 \otimes u_6) = \kappa_7^2 (w_7 \otimes 1), \tag{16.23}$$

et que $\kappa_7^2(w_{10} \otimes u_6)$ est un d_7-cocycle, non nul, qui engendre $E_7^{10,6}$. Mais E_∞ est triviale, et $d_r E_r^{10,6} = 0$ pour tout $r > 7$, il doit donc exister un indice $t \geqq 7$ et un élément z tels que:

$$0 \neq \kappa_t^2 (w_{10} \otimes u_6) = d_t(z), \qquad (z \varepsilon E_t^{10-t, 5+t}) ; \tag{16. 24}$$

cependant, $E_r^{p,q} = 0$ pour $0 < p \leqq 3$, l'égalité (16. 24) n'est donc possible que pour $t = 10$ et

$$z \varepsilon E_{10}^{0,15} \subset E_7^{0,15}.$$

$E_7^{0,15}$ est, d'après (16. 22), de dimension 2, engendré par $u_6 \cdot u_9$ et u''_{15} ; comme:

$$d_7 \kappa_7^2 (1 \otimes u_6 \cdot u_9) = \kappa_7^2 (w_7 \otimes u_9)$$

est $\neq 0$ d'après (16. 22), on doit avoir

$$z = u''_{15} + c\, u_6 \cdot u_9, \qquad (c \varepsilon Z_2).$$

et $z = u_{15}$ est un élément congru à u^*_{15} mod. D qui vérifie 16. 2(a) ; l'assertion 16. 2(b), résulte de (16. 22) ; enfin, $E_7^{10,0}$ est, vu (16. 22), engendré par $\kappa_7^2(w_{10} \otimes 1)$, d'où

$$0 \neq d_{10} \kappa_{10}^2 (1 \otimes u_9) = \kappa_{10}^2 (w_{10} \otimes 1)$$

ce qui prouve 16. 2(c).

Notations. h^* et h_* seront les homomorphismes induits en cohomologie et en homologie mod. 2 par le produit de $\boldsymbol{Spin}(10)$.

$\{v_{i_1,\ldots,i_k}\}$ sera la base de $H_*(\boldsymbol{Spin}(10), Z_2)$, duale de la base $\{u_{i_1} \cdot \cdots u_{i_k}\}$ de $H^*(\boldsymbol{Spin}(10), Z_2)$, $(i_1 < \cdots < i_k ; 1 \leqq k \leqq 6)$.

$H_*(\boldsymbol{Spin}(10), Z_2)$ opère sur $H^*(\boldsymbol{Spin}(10), Z_2)$, ainsi que sur l'algèbre spectrale universelle, (cf. No. 3), et nous noterons $\vartheta_{i_1,\ldots,i_k}$ l'opérateur défini par $v_{i_1,\ldots,i_k}$.

LEMME 16.3. *On a* $h^*(u_i) = u_i \otimes 1 + 1 \otimes u_i$, $(i < 15)$, *et*

$$h^*(u_{15}) = u_{15} \otimes 1 + 1 \otimes u_{15} + u_9 \otimes u_6.$$

Pour u_i, $(i < 15)$, il suffit le remarquer qu'un élément universellement transgressif est aussi primitif, ([3], Prop. 20.1).

En utilisant les propriétés des opérateurs ϑ_u énoncées au No. 13, et (16.2), on obtient:

$$d_r\kappa_r{}^2(1 \otimes u_{15} \cdot \vartheta_{i_1,\ldots,i_k}) = d_r\kappa_r{}^2(1 \otimes u_{15})) \cdot \vartheta_{i_1,\ldots,i_k}, \qquad (r \leqq 9).$$

$$(16.31) \qquad d_{10}\kappa_{10}{}^2(1 \otimes u_{15} \cdot \vartheta_6) = (d_{10}\kappa_{10}{}^2(1 \otimes u_{15})) \cdot \vartheta_6 = (\kappa_{10}{}^2(w_{10} \otimes u_6)) \cdot \vartheta_6,$$

$$d_{10}\kappa_{10}{}^2(1 \otimes u_{15} \cdot \vartheta_6) = \kappa_{10}{}^2(w_{10} \otimes u_6 \cdot \vartheta_6) = \kappa_{10}{}^2(w_{10} \otimes 1) \neq 0,$$

donc

$$(16.32) \qquad u_{15} \cdot \vartheta_6 \neq 0.$$

Par ailleurs, 16.2(b) montre que $H^*(\boldsymbol{Spin}(10), Z_2)$ ne contient qu'un élément non nul de degré strictement positif et < 15 qui soit un d_r-cocycle pour $2 \leqq r \leqq 9$, c'est u_9; les formules (16.31) et (16.32) entraînent par conséquent:

$$u_{15} \cdot \vartheta_6 = u_9$$

$$u_{15} \cdot \vartheta_{i_1,\ldots,i_k} = 0, \qquad (i_1 + \cdots + i_k < 15, (i_1, \cdots, i_k) \neq (6)).$$

Si l'on remonte à la définition de ϑ_u, No. 3.1, on voit que cela signifie précisément que $h^*(u_{15})$ a la valeur annoncée dans (16.3).

THÉORÈME 16.4. *$H_*(\boldsymbol{Spin}(10), Z_2)$ possède un système simple de générateurs $v_3, v_5, v_6, v_7, v_9, v_{15}$, $(Dv_i = i)$, vérifiant les relations:*

$$v_i \vee v_i = 0, \quad (3 \leqq i \leqq 15), \qquad v_i \vee v_j = v_j \vee v_i \quad (i < j, (i, j) \neq (6, 9)),$$

$$v_6 \vee v_9 = v_9 \vee v_6 + v_{15}.$$

On tire tout d'abord de (16.3):

(16.41). h^* est symétrique (au sens de la définition 2.3) sur $\Delta(u_3, u_5, u_6, u_7, u_9)$.

(16.42). $h^*(a \cdot u_{15}) = h^*(a)(u_{15} \otimes 1 + 1 \otimes u_{15}) + h^*(a) \cdot (u_9 \otimes u_6)$, et le premier addende du deuxième membre est symétrique si $h^*(a)$ l'est. Il est clair que

$$\langle h^*(a) \cdot (u_9 \otimes u_6), v_{i_1} \cdot \cdots \cdot v_{i_j} \otimes v_{i_{j+1}} \rangle = 0 \text{ si } i_1 < \cdots < i_j < i_{j+1},$$

$$\langle h^*(a) \cdot (u_9 \otimes u_6), v_i \otimes v_j \rangle = 0 \text{ si } (i, j) \neq (9, 6).$$

ce qui, joint à (16.41) et (16.42) montre que

$$v_i \vee v_j = v_j \vee v_i \text{ pour } (i < j, (i, j) \neq (6, 9)),$$

et, par récurrence sur j, que

$$v_{i_1, \ldots, i_j} = v_{i_1} \vee v_{i_2} \vee \cdots \vee v_{i_j}, \qquad (i_1 < i_2 < \cdots < i_j)$$

donc en particulier que (v_i) est un système simple de générateurs.

De (16.3) et (16.42) on tire que

$$h^*(u_{i_1} \cdot \cdots \cdot u_{i_k}), \qquad (i_1 < i_2 < \cdots < i_k),$$

ne contient jamais de terme de la forme $u_i \otimes u_i$, d'où les égalités $v_i \vee v_i = 0$.

Enfin $H^{15}(\boldsymbol{Spin}(10), Z_2)$ est engendré par u_{15} et par

$$D^{15} = H^{15}(\boldsymbol{Spin}(10), Z_2) \cap \Delta(u_3, u_5, u_6, u_7, u_9);$$

par définition des v_i on a $\langle D^{15}, v_{15} \rangle = 0$, et, vu (16.3), on a

$$\langle x, v_6 \vee v_9 \rangle = \langle x, v_9 \vee v_6 \rangle \qquad (x \in D^{15})$$

$$\langle u_{15}, v_9 \vee v_6 \rangle = \langle u_{15}, v_{15} \rangle = 1$$

$$\langle u_{15}, v_6 \vee v_9 \rangle = 0$$

ce qui implique

$$v_6 \vee v_9 = v_9 \vee v_6 + v_{15}.$$

Remarque. Nous ne pousserons pas plus avant l'étude de $H_*(\boldsymbol{Spin}(n), Z_2)$, et nous bornerons à signaler qu'en s'appuyant sur les résultats des Nos. 15, 16, et sur un théorème non encore publié de H. Cartan on peut voir que cette algèbre n'est pas commutative pour $n \geqq 10$.

Principe de démonstration. Nous reprenons les notations de 12.1(c) et soit $\{v_{i_1, \ldots, i_k}\}$ une base de $H_*(\boldsymbol{Spin}(n), Z_2)$, duale de la base formée par les monômes basiques associés au système simple (u_i, u), $(1 < i < n - s(n)$, v_i correspondant à u_i, $(1 \leqq i < n - s(n))$, et $v_{n-s(n)}$ à u. Les u_i étant primitifs, on voit tout d'abord aisément que $v_i \vee v_i = 0$ $(1 < i < n - s(n))$; si maintenant $H_*(\boldsymbol{Spin}(n), Z_2)$ était commutative, l'application diagonale en ferait

une algèbre de Hopf, à laquelle on pourrait appliquer les résultats du No. 1; les éléments v_i, $(1 < i < n - s(n))$, feraient alors partie d'un système de type M lequel le dernier générateur $v^* \equiv v_{n-s(n)}$ mod. D serait de carré nul, (même argument que dans 14(iii)), et ainsi $H_*(\boldsymbol{Spin}(n), Z_2)$ serait isomorphe à $\wedge(v_1, \cdots, v_{n-s(n)})$. Mais alors, d'après H. Cartan, $H^*(B_{\boldsymbol{Spin}(n)}, Z_2)$ aurait même série de Poincaré qu'une algèbre de polynômes à $n - s(n)$ variables de degrés $Du_i + 1$, $Du + 1$, ce qui contredit (15. 5) pour $n \geqq 10$.

V. Les deux premiers groupes exceptionnels.

17. Le groupe G_2. On sait que le groupe compact $\boldsymbol{G}_2$ peut être envisagé comme le groupe des automorphismes des octaves de Cayley; il est de rang 2 et possède 14 paramètres.

LEMME 17. 1. *$\boldsymbol{G}_2$ contient un sous-groupe isomorphe à $\boldsymbol{SU}(2)$, tel que l'espace homogène $\boldsymbol{G}_2/\boldsymbol{SU}(2)$ soit homéomorphe à la variété de Stiefel $\boldsymbol{V}_{7,2}$ des 2-repères orthonormaux de R^7.*

Soient (e_i), $(i \in Z_7)$, une base des octaves purement imaginaires, les systèmes quaternioniens étant les triplets (e_i, e_{i+1}, e_{i+3}). Soient f_1 et f_2 deux octaves purement imaginaires orthonormaux, relativement à la métrique usuelle; posons $f_4 = f_1 \cdot f_2$ et soit f_3 orthonormal à f_1, f_2, f_4; On voit sans difficulté que

$$f_5 = f_2 \cdot f_3, \qquad f_6 = f_3 \cdot f_4, \qquad f_7 = f_4 \cdot f_5,$$

forment avec les f_i, $(1 \leqq i \leqq 4)$, un système libre et que la correspondance $e_i \to f_i$, $(i \in Z_7)$ définit un automorphisme des octaves, (pour plus de détails, *voir* N. Jacobson, *Duke Mathematical Journal*, 5 (1939), pp. 776-783). Cela montre:

(i) $\boldsymbol{G}_2$ est transitif sur les systèmes f_1, f_2 de deux octaves purement imaginaires orthonormaux, autrement dit sur $\boldsymbol{V}_{7,2}$.

(ii). Les éléments de $\boldsymbol{G}_2$ qui laissent f_1 et f_2 fixes sont en correspondance biunivoque avec les vecteurs unitaires f_3 de l'espace à 4 dimensions orthogonal à f_1, f_2, f_4, c'est à dire avec les points de $\boldsymbol{S}_3$; le groupe d'isotropie relativement à $\boldsymbol{V}_{7,2}$ est donc homéomorphe à $\boldsymbol{S}_3$, il est bien isomorphe à $\boldsymbol{SU}(2)$.

THÉORÈME 17. 2. *$H^*(\boldsymbol{G}_2, Z)$ possède un système de 2 générateurs h_3, h_{11} $(Dh_i = i)$, jouissant des propriétés suivantes:*

(a). $h_3^4 = h_{11}^2 = h_3^2 \cdot h_{11} = 0$.

(b). *$H^*(\boldsymbol{G}_2, Z)$ est la somme directe faible des sous-groupes cycliques engendrés par les éléments d'ordre infini $h_3, h_{11}, h_3 \cdot h_{11}$ et par les éléments d'ordre deux h_3^2, h_3^3.*

Compte tenu de la structure de $H^*(\boldsymbol{V}_{7,2}, Z)$, rappelée au No. 8, ce théorème résulte visiblement du:

THÉORÈME 17.3. (a). *$H^*(\boldsymbol{S}_3 \times \boldsymbol{V}_{7,2}, Z)$ est l'algèbre graduée associée à $H^*(\boldsymbol{G}_2, Z)$ convenablement filtrée.*

(b). Pour $p \neq 2$, $H^*(\boldsymbol{G}_2, Z_p) \cong \wedge(x_3, x_{11})$, $(Dx_i = i)$; *les puissances réduites y sont triviales si $p \neq 5$ et $\mathcal{P}_5^1(x_3) = x_{11}$.*[11]

(c). *$H^*(\boldsymbol{G}_2, Z_2)$ possède un système simple de générateurs universellement transgressifs x_3, x_5, x_6, $(Dx_i = i)$, liés par les relations:*

$$Sq^2 x_3 = x_5,\ Sq^1 x_5 = x_6,\ Sq^3 x_3 = x_3,\ Sq^i x_j = 0 \text{ sinon.}$$

Dans la suite spectrale de la fibration $(\boldsymbol{G}_2, \boldsymbol{V}_{7,2}, \boldsymbol{S}_3)$ on a pour $\Gamma = Z$ ou Z_p

$$E_2 \overset{\bullet}{\cong} H^*(\boldsymbol{V}_{7,2}, \Gamma) \otimes H^*(\boldsymbol{S}_3, \Gamma),$$

($\otimes$ sur Γ), d'où $E_2 = E_\infty$ pour des raisons de degrés évidentes, ce qui établit 17.3(a).

Considérons maintenant la fibration $\boldsymbol{Spin}(7)/\boldsymbol{G}_2 = \boldsymbol{S}_7$, (*voir* [1]), et soit i l'inclusion d'une fibre. D'après le No. 8, 12.1(a) et 17.3(a), on a pour les polynômes de Poincaré:

$$P_p(\boldsymbol{S}_7, t) \cdot P_p(\boldsymbol{G}_2, t) = P_p(\boldsymbol{Spin}(7), t), \qquad (p \text{ nul ou premier}),$$

et ainsi $\boldsymbol{G}_2$ est totalement non homologue à zéro mod. p dans $\boldsymbol{Spin}(7)$,[12] en particulier, $H^*(\boldsymbol{G}_2, Z_2)$ admettra comme système simple de générateurs les éléments $i^*(u_i)$, (notations de 12.1(c), $(1 \leqq i \leqq 3)$), d'où les Sq^i de 17.3(c); l'élément de plus petit degré positif $x_3 = i^*(u_1)$ étant forcément universellement transgressif, il en sera de même pour

$$x_5 = Sq^2 x_3 \text{ et } x_6 = Sq^3 x_3$$

ce qui termine la démonstration de 17.3(c).

Enfin, mod. p impair, $H^*(\boldsymbol{Spin}(7), Z_p)$ s'identifie canoniquement à $H^*(\boldsymbol{SO}(7), Z_p)$, possède trois générateurs b_1, b_2, b_3 de degrés 3, 7, 11 et

---

[11] Pour un choix convenable de générateurs universellement transgressifs x_i.

[12] Il l'est aussi relativement à la cohomologie entière, comme on peut le voir en montrant que l'algèbre spectrale sur Z de la fibration $(\boldsymbol{Spin}(7), \boldsymbol{S}_7, \boldsymbol{G}_2)$ est triviale; cela résultera aussi de la remarque du No 23.

$H^*(\boldsymbol{G}_2, Z_p)$ est l'algèbre engendrée par $i^*(b_1)$ et $i^*(b_3)$, et 17.3(b) résulte de 8.9.

THÉORÈME 17.4. *Pour $p \neq 2$, $H_*(\boldsymbol{G}_2, Z_p)$ est une algèbre extérieure anticommutative, isomorphe à $H^*(\boldsymbol{G}_2, Z_p)$.*

$H_*(\boldsymbol{G}_2, Z_2)$ est une algèbre extérieure commutative, engendrée par 3 éléments de degrés 3, 5, 6.

C'est une conséquence de (2.8) et (17.3).

18. Groupes opérant sur un espace fibré principal. Nous aurons besoin au No 20 de quelques remarques relatives à une situation souvent considérée par J. Leray (*voir* par ex. [14]) et aussi dans [3], aussi pour plus de clarté, voulons-nous énoncer ici, sous des hypothèses générales, les propriétés en question.

Soient (E, B, G, q) un espace fibré principal connexe, de groupe structural G, compact pour fixer les idées, U un sous-groupe fermé de G et N un sous-groupe fermé du normalisateur de U dans G. Les éléments de N définissent des homéomorphismes de E compatibles avec la fibration $(E, E/U, U, s)$, (i. e. ils envoient fibres sur fibres) ; par conséquent N opère sur les algèbres de cohomologie de ces espaces et sur l'algèbre spectrale de cette fibration. Supposons dorénavant U connexe. On a alors les propriétés suivantes:

(a). *Par l'intermédiaire des translations, $N/N \cap U$ opère sur*

$$H^*(E/U, \Gamma), \qquad H^*(E, \Gamma),$$

en commutant à s^* et en respectant la filtration de $H^*(E, \Gamma)$, (Γ, anneau de coefficients).

(b). *$N/N \cap U$ opère sur l'algèbre spectrale (E_r) de $(E, E/U, U, s)$ pour $r \geqq 2$ [13] ; il commute avec l'isomorphisme canonique de $E_2^{p,0}$ sur $H^p(E/U, \Gamma)$; sur $E_2^{0,q}$ identifié canoniquement à $H^q(U, \Gamma)$, il opère par l'intermédiaire des automorphismes $u \to n^{-1} \cdot u \cdot n$, $(u \in U, n \in N)$.*

(c). *Si Γ est un corps, il existe un isomorphisme d'espaces vectoriels gradués de E_∞ sur $H^*(E, \Gamma)$ commutant à $N/N \cap U$.*[14]

Ces propriétés, qui sont bien naturelles, se démontrent aisément si l'on

---

[13] Cela signifie qu'à tout élément de $N/N \cap U$ correspond un endomorphisme (de la structure d'algèbre bigraduée) de E_r, $(r \geqq 2)$, commutant à d_r et à $\kappa_{r+1}{}^r$, cette correspondance étant homomorphique. N opère sur toute l'algèbre spectrale, mais $N \cap U$ n'opère trivialement que pour $r \geqq 2$, d'où cette restriction.

[14] Pour E_∞ il s'agit bien entendu du degré total.

remonte à la définition et aux principales propriétés de l'algèbre spectrale des espaces fibrés (résumées par exemple dans [3], § 4), et si l'on utilise pour définir cette dernière des couvertures fines de E et de E/U sur lesquelles N opère, (par exemple les cochaînes d'Alexander-Spanier).

N agit aussi sur la fibration $(E/U, E/G, G/U, q)$ en laissant chaque fibre invariante, et sur G/U envisagé comme espace des classes de restes à gauche, il opère par les translations à droite; ces opérations jouissent des propriétés suivantes, analogues aux précédentes:

(d). *$N/N \cap U$ opère sur $H^*(E, \Gamma)$ en respectant la filtration, opère trivialement sur $H^*(E/G, \Gamma)$, et commute à q^*.*

(e). *Supposons G connexe: Alors $N/N \cap U$ opère sur l'algèbre spectrale (E_r) $(r \geqq 2)$,[13] trivialement sur $E_2^{p,0}$, et commute avec l'isomorphisme canonique de $E_2^{0,q}$ sur $H^q(G/U, \Gamma)$.*

(f). *Si Γ est un corps il existe un isomorphisme d'espaces vectoriels gradués de E_∞ sur $H^*(E/U, \Gamma)$ commutant à $N/N \cap U$.*[14]

19. Homologie et cohomologie de F_4, (résultats). $\boldsymbol{F_4}$ est de rang 4, à 52 paramètres, il agit transitivement sur le " plan projectif des octaves " $\boldsymbol{W}$, variété à 16 dimensions, et le groupe d'isotropie est isomorphe à $\boldsymbol{Spin}(9)$, (voir [1]); dans $\boldsymbol{W}$ le complément d'une droite projective, (i. e. d'une sphère à 8 dimensions), est homéomorphe à R^{16}, d'où

$$H^*(\boldsymbol{W}, Z) \cong Z[x]/(x^3), \qquad (Dx = 8). \tag{19.1}$$

THÉORÈME 19.2. (a). *$\boldsymbol{F_4}$ n'a pas de p-torsion pour $p \neq 2, 3$ et*

$$H^*(\boldsymbol{F_4}, Z_p) = \wedge(x_3, x_{11}, x_{15}, x_{23}), \qquad (Dx_i = i).$$

(b). *$\boldsymbol{F_4}$ possède de la 3-torsion et*

$$H^*(\boldsymbol{F_4}, Z_3) \cong Z_3[x]/(x^3) \otimes \wedge(x_3, x_7, x_{11}, x_{15}), \qquad (Dx = 8, Dx_i = i).$$

(c). *$\boldsymbol{F_4}$ a de la 2-torsion; $H^*(\boldsymbol{F_4}, Z_2)$ possède un système simple de générateurs universellement transgressifs $x_3, x_5, x_6, x_{15}, x_{23}$ et $H^*(\boldsymbol{F_4}, Z_2) \cong H^*(\boldsymbol{G_2} \times \boldsymbol{S_{15}} \times \boldsymbol{S_{23}}, Z_2)$ au sens du cup-produit et, pour $D \leqq 22$, au sens des Sq^i.*

(d). *$H^*(\boldsymbol{F_4}, Z)$ est additivement et multiplicativement isomorphe au produit tensoriel $U \otimes H^*(\boldsymbol{G_2} \times \boldsymbol{S_{15}}, Z)$, où U est un anneau unitaire gradué dans lequel $U^0 = U^{23} = Z$, $U^8 \cong U^8 \cdot U^8 = U^{16} \cong Z_3$, $U^i = 0$ sinon; ses coefficients de torsion sont donc égaux à 2, 3, ou 6.*

Remarques.

19. 3. Les isomorphismes 19. 2(a) et (b) sont bien entendu valables pour la structure additive et le cup-produit; nous obtiendrons également les renseignements suivants au sujet des puissances réduites:

(a). $\mathcal{P}_5{}^1 x_3 = x_{11}$ $\mathcal{P}_7{}^1 x_3 = x_{15}$.[11]

(b). $\mathcal{P}_3{}^1 x_3 = x_7$, $\mathcal{P}_3{}^1 x_{11} = x_{15}$, *et* x (*notations de* 19. 2(b)), *provient de* x_7 *par l'homomorphisme de Bockstein* (*pour un choix convenable de* x_i, x).

19. 4. Pour $p \geqq 5$, on peut supposer les x_i universellement transgressifs; on a donc toujours

$$\mathcal{P}_p{}^i x_j = a_{p,j}{}^i x_k, \qquad (a_{p,j}{}^i \in Z_p,\ k = 2i(p-1) + j),$$

et en particulier $\mathcal{P}_p{}^i$ est identiquement nulle sur $H^*(F_4, Z_p)$ pour $p \geqq 13$, $i \geqq 1$.

19. 5. Nous ignorons si l'isomorphisme de 19. 2(c) est valable pour les i-carrés en général; en fait, comme nous le verrons, la seule question qui reste à trancher est de savoir si $Sq^8 x_{15}$ est nul ou égal à x_{23}.

THÉORÈME 19. 6. *Pour* $p \neq 2, 3$ $H_*(\boldsymbol{F}_4, Z_p)$ *est une algèbre extérieure anticommutative, isomorphe à* $H^*(\boldsymbol{F}_4, Z_p)$.

$H_*(\boldsymbol{F}_4, Z_2)$ *est une algèbre extérieure commutative engendrée par* 5 *éléments de degrés* 3, 5, 6, 15, 23.

Cela résulte de (2. 8) et (19. 2). Le reste de ce travail est consacré à la démonstration de 19. 2, 19. 3 et 19. 5.

20. L'homomorphisme Ψ^*: $H^8(\boldsymbol{F}_4/\boldsymbol{Spin}(9), \boldsymbol{Z}) \to H^8(\boldsymbol{F}_4, \boldsymbol{Z})$. Nous choisissons une fois pour toutes une suite de sous-groupes emboîtés

$$\boldsymbol{F}_4 \supset \boldsymbol{Spin}(9) \supset \boldsymbol{Spin}(8) \supset \boldsymbol{T}$$

où $\boldsymbol{T}$ est un tore à quatre dimensions, maximal dans chacun des trois autres groupes. Nous considérerons les projections naturelles

$$\boldsymbol{F}_4 \xrightarrow{\alpha} \boldsymbol{F}_4/\boldsymbol{Spin}(8) \xrightarrow{\beta} \boldsymbol{F}_4/\boldsymbol{Spin}(9) = \boldsymbol{W} \qquad \text{et } \psi = \beta \circ \alpha$$

Soient N_G le normalisateur de T dans G et $\Phi(G) = N_G/T$, ($G = \boldsymbol{F}_4$, $\boldsymbol{Spin}(9)$, $\boldsymbol{Spin}(8)$). Ce dernier est le groupe de Weyl de G, qui s'identifie au groupe des automorphismes de T (ou du revêtement universel de T), qui sont induits par les automorphismes intérieurs de G laissant T invariant.

LEMME 20. 1. *$\Phi(\boldsymbol{Spin}(8))$ est invariant dans $\Phi(F_4)$ et $\Phi(F_4)/\Phi(\boldsymbol{Spin}(8))$ est isomorphe au groupe σ_3 des permutations de 3 objets; N_{F_4} fait partie du normalisateur de $\boldsymbol{Spin}(8)$.*

Bien que nous ne puissions guère indiquer de référence précise, ce lemme est connu et résulte trivialement de propriétés classiques de F_4, et nous en donnerons la démonstration sans entrer dans tous les détails.[15]

Pour un choix convenable des coordonnées x_i dans le revêtement universel R^4 de T, les racines de F_4 s'écrivent, en omettant le facteur $2\pi\sqrt{-1}$:

$$\pm x_i \pm x_j \qquad (i<j), \tag{20.2}$$

$$1/2(\pm x_1 \pm x_2 \pm x_3 \pm x_4)\,;\ \pm x_i, \qquad (1 \leqq i \leqq 4), \tag{20.3}$$

où (20. 2) est l'ensemble des racines de $\boldsymbol{Spin}(8)$. Soit S^*_α la transformation contragrédiente de la symétrie à l'hyperplan $\alpha = 0$ [16]; de la formule

$$S^*_\alpha(\beta) = \beta - 2(\alpha,\beta)\cdot(\alpha,\alpha)^{-1}\alpha, \qquad (\alpha,\beta \text{ formes linéaires sur } R^4),$$

où (,) est le produit scalaire défini par la métrique,[16] on déduit immédiatement que le système (20. 2) est invariant par les symétries S^*_α, où α parcourt les racines de F_4; comme le groupe de Weyl d'un groupe de Lie compact connexe G est engendré par les symétries aux hyperplans obtenus en égalant à zéro les racines de G, il en résulte que $\Phi(\boldsymbol{Spin}(8))$ est invariant dans $\Phi(F_4)$.

$\Phi(F_4)$ et $\Phi(\boldsymbol{Spin}(8))$ sont d'ordres respectifs 192. 6 et 192; leur quotient est bien d'ordre 6; on vérifie directement, ce que nous ne ferons pas, qu'il est isomorphe à σ_3 (en fait au groupe des permutations de trois des sommets d'un simplexe fondamental de $\boldsymbol{Spin}(8)$); on peut du reste s'épargner cette verification en s'appuyant sur des résultats connus: En effet, nous avons vu que $\Phi(F_4)/\Phi(\boldsymbol{Spin}(8))$ est un sous-groupe d'ordre 6 du quotient $\Phi'(\boldsymbol{Spin}(8))/\Phi(\boldsymbol{Spin}(8))$, où $\Phi'(\boldsymbol{Spin}(8))$ est le groupe de Cartan de $\boldsymbol{Spin}(8)$, c'est à dire le groupe de tous les automorphismes de R^4 laissant le système des racines (20. 2) invariant, et l'on sait que ce quotient est isomorphe à σ_3.[17]

---

[15] Pour tous les résultats de la théorie des groupes de Lie utilisés ici, *voir* par exemple E. Cartan, loc. cit., 0. 7, F. Gantmacher, Rec. Math. Moscou N. S. *5* (1939), 101-144, E. Stiefel, Comm. Math. Helv. *14* (1941-42), pp. 350-380.

[16] Au sens de la métrique définie par la somme des carrés des racines de F_4.

[17] Ce n'est qu'une autre manière de dire que la quotient du groupe des automorphismes de $\boldsymbol{Spin}(8)$ par le sous-groupe des automorphismes intérieurs est isomorphe à σ_3, (principe de trialité); en somme, (20. 1) montre que tous ces automorphismes sont induits par des automorphismes intérieurs de F_4.

Enfin, les automorphismes $t \to n \cdot t \cdot n^{-1}$ ($t \in T$, $n \in \boldsymbol{N}_{\boldsymbol{F}_4}$) laissant le système (20.2) invariant, il est clair que $\boldsymbol{N}_{\boldsymbol{F}_4}$ est dans le normalisateur de $\boldsymbol{Spin}(8)$.

LEMME 20.4. *Au point de vue additif,*

$$H^*(\boldsymbol{F}_4/\boldsymbol{Spin}(8), Z) \cong H^*(\boldsymbol{W}, Z) \otimes H^*(\boldsymbol{S}_8, Z).$$

En effet, $\boldsymbol{Spin}(9)/\boldsymbol{Spin}(8) = \boldsymbol{S}_8$, le terme

$$E_2 \cong H^*(\boldsymbol{W}, Z) \otimes H^*(\boldsymbol{S}_8, Z)$$

de la suite spectrale de la fibration $(\boldsymbol{F}_4/\boldsymbol{Spin}(8), \boldsymbol{W}, \boldsymbol{S}_8)$ ne contient que des éléments de degrés pairs et est donc égal à E_∞.

LEMME 20.5. *La représentation de* $\sigma_3 = \boldsymbol{N}_{\boldsymbol{F}_4}/\boldsymbol{N}_{\boldsymbol{F}_4} \cap \boldsymbol{Spin}(8)$ *dans* $H^8(\boldsymbol{F}_4/\boldsymbol{Spin}(8), Z)$ *induite par les translations à droite est fidèle.*

Il suffira évidemment de montrer que la représentation correspondante dans $H^8(\boldsymbol{F}_4/\boldsymbol{Spin}(8), R)$ est fidèle (R corps des réels).

Relativement à R, $\boldsymbol{Spin}(8)/\boldsymbol{T}$ est totalement non homologue à zéro dans la fibration $(\boldsymbol{F}_4/\boldsymbol{T}, \boldsymbol{F}_4/\boldsymbol{Spin}(8), \boldsymbol{Spin}(8)/\boldsymbol{T})$, ([3], Prop. 26.1 et Cor. à la Prop. 18.3), donc

$$E_\infty \cong E_2 \cong H^*(\boldsymbol{F}_4/\boldsymbol{Spin}(8), R) \otimes H^*(\boldsymbol{Spin}(8), R),$$

(ce qui du reste résulte aussi du fait que E_2 ne contient que des éléments de degrés pairs d'après la formule de Hirsch).

Par ailleurs, si T est un tore maximal d'un groupe de Lie compact connexe G, la représentation naturelle de $\Phi(G) = N_T/T$ dans $H^*(G/T, R)$ est équivalente à la représentation régulière d'après J. Leray ([14], Prop. 11.1, ou [3], Lemme 27.1). En appliquant les remarques du No. 18 au cas

$$E = G = \boldsymbol{F}_4, \qquad N = \boldsymbol{N}_{\boldsymbol{F}_4}, \qquad U = \boldsymbol{T},$$

on voit donc que E_∞ (resp. $1 \otimes H^*(\boldsymbol{Spin}(8)/\boldsymbol{T}, R)$), est espace de la représentation régulière de $\Phi(\boldsymbol{F}_4)$, (resp. $\Phi(\boldsymbol{Spin}(8))$), et que $\Phi(\boldsymbol{Spin}(8))$ agit trivialement sur $H^*(\boldsymbol{F}_4/\boldsymbol{Spin}(8), R) \otimes 1$. Il s'ensuit évidemment que la représentation de $\sigma_3 = \Phi(\boldsymbol{F}_4)/\Phi(\boldsymbol{Spin}(8))$ induite dans ce dernier espace est la représentation régulière, et les 2 sous-espaces invariants de dimension 2 $H^8(\boldsymbol{F}/\boldsymbol{Spin}(8), R)$, $H^{16}(\boldsymbol{FSpin}(8), R)$ sont espaces de la représentation irréductible non triviale de σ_3, qui est fidèle comme on sait.

LEMME 20.6. *On a* $\alpha^*(3x)$ *pour tout* $x \in H^8(\boldsymbol{F}_4/\boldsymbol{Spin}(8), Z)$.

$\Phi(\boldsymbol{Spin}(8))$ opère trivialement sur $H^*(\boldsymbol{F}_4, Z)$ (puisque ces opérations

sont induites par des translations homotopes à l'identité), fidèlement sur $H^8(\boldsymbol{F_4/Spin}(8), Z)$ et commute à α^*; soit v un élément d'ordre 3 de σ_3, et soit A l'automorphisme de $H^8(\boldsymbol{F_4/Spin}(8), R)$ ou de $H^8(\boldsymbol{F_4/Spin}(8), Z)$ qui lui correspond; cet espace étant de dimension 2 et la représentation étant fidèle, l'automorphisme A n'a pas de valeur propre réelle, et l'égalité

$$(A^2 + A + 1) \cdot (A - 1) = A^3 - 1 = 0$$

entraîne

$$A^2 + A + 1 = 0$$

$$\alpha^*(A^2(x) + A(x) + x) = 0$$

et le lemme, puisque

$$\alpha^*(A(x)) = A(\alpha^*(x)) = \alpha^*(x).$$

PROPOSITION 20.7. *L'image de* $\psi^*: H^8(\boldsymbol{W}, Z) \to H^8(\boldsymbol{F_4}, Z)$ *est égale à* $H^8(\boldsymbol{F_4}, Z)$; *elle est soit nulle, soit isomorphe à* Z_3.

On obtient la seconde assertion en appliquant (20.6) à $\beta^*(x)$, où x engendre $H^8(\boldsymbol{W}, Z)$. Pour établir la première il suffit de remarquer que dans l'algèbre spectrale sur Z de la fibration $(\boldsymbol{F_4}, \boldsymbol{W}, \boldsymbol{Spin}(9), \psi)$, on a, vu (12.1):

$${}^8E_2 = E_2^{8,0} \cong H^8(\boldsymbol{W}, Z)$$

d'où

$${}^8E_\infty = E_\infty^{8,0} \text{ et } \psi^*(H^8(\boldsymbol{W}, Z)) = H^8(\boldsymbol{F_4}, Z).$$

21. Cohomologie mod. p de $\boldsymbol{F_4}$, $(p \neq 2)$. Nous voulons démontrer ici 19.2(a), (b) et 19.3, en déterminant l'algèbre spectrale mod. p de $(\boldsymbol{F_4}, \boldsymbol{W}, \boldsymbol{Spin}(9), \psi)$. On a

$$E_2 \cong H^*(\boldsymbol{W}, Z_p) \otimes H^*(\boldsymbol{Spin}(9), Z_p) \cong Z_p[x]/(x^3) \otimes \wedge(b_1, b_2, b_3, b_4)$$

$(Dx = 8,\ Db_i = 4i - 1,\ b_i$ universellement transgressif). Evidemment, $E_2 = E_8$.

Soit $p \neq 2, 3$; on tire de (20.7) que $\psi^*(3x) = 0$, ce qui n'est possible que si

$$d_8\kappa_8^2(1 \otimes b_2) = \kappa_8^2(cx \otimes 1),\ (c \in Z_p, c \neq 0);$$

les d_8-cocycles forment donc la sous-algèbre

$$H^*(\boldsymbol{W}, Z_p) \otimes \wedge(b_1, b_3, b_4) + x^2 \otimes b_2 \otimes \wedge(b_1, b_3, b_4)$$

et

$$d_8(E_8) = (x) \otimes \wedge(b_1, b_3, b_4);$$

ainsi, E_9 est l'algèbre extérieure engendrée par les éléments $\kappa_9^2(x^2 \otimes b_2)$ et $\kappa_9^2(1 \otimes b_i)$, $(i = 1, 3, 4)$, de degrés 23, 3, 11, 15. Les différentielles d_r, $(r \geqq 9)$ sont nulles sur les deux premiers; d'autre part on voit que

$$E_9^{12,0} = E_9^{16,0} = 0$$

et les différentielles d_r $(r \geqq 2)$ seront donc toutes nulles sur les éléments transgressifs $1 \otimes b_3$ et $1 \otimes b_4$; cela montre que $d_r = 0$ $(r \geqq 9)$, d'où (compte tenu de la Prop. 8.1 de [3]):

$$\wedge(x_3, x_{11}, x_{15}, x_{23}) \cong E_9 \cong E_\infty \cong H^*(\boldsymbol{F}_4, Z_p).$$

Les polynômes de Poincaré de $\boldsymbol{F}_4$ en caractéristique zéro et mod. p $(p \geqq 5)$, sont donc égaux, d'où l'absence de p-torsion et 19.2(a).

De plus, on peut prendre comme générateurs x_3, x_{11}, x_{15} les éléments univoquement déterminés tels que

$$i^*(x_{4j-1}) = b_j, \qquad (j = 1, 3, 4,\ i \text{ inclusion de } \boldsymbol{Spin}(9) \text{ dans } \boldsymbol{F}_4),$$

et les formules

$$\mathcal{P}_5^1 x_3 = x_{11}, \qquad \mathcal{P}_7^1 x_3 = x_{15},$$

résultent de (8.9).

Soit maintenant $p = 3$. L'algèbre $H^*(\boldsymbol{W}, Z)$ ne contenant pas d'élément non nul de degré 4 ou 12, les éléments universellement transgressifs b_1, b_3 sont des cocycles pour toutes les différentielles, autrement dit font partie de l'image de i^*; il en sera alors de même pour

$$b_2 = \mathcal{P}_3^1 b_1 \text{ et } b_4 = \mathcal{P}_3^1 b_3$$

(voir (8.9)), $\boldsymbol{Spin}(9)$ est totalement non homologue à zéro mod. 3 dans $\boldsymbol{F}_4$, d'où, compte tenu de (7.4):

$$H^*(\boldsymbol{F}_4, Z_3) \cong E_\infty \cong E_2 \cong H^*(\boldsymbol{W}, Z_3) \otimes H^*(\boldsymbol{Spin}(9), Z_3)$$

ce qui établit 19.2(b); on voit en outre que l'on peut prendre comme système 3-semi-libre de générateurs de $H^*(\boldsymbol{F}_4, Z_3)$ des éléments x_i tels que

$$x_8 = \psi^*(x), \qquad i^*(x_{4j-1}) = b_j, \qquad (1 \leqq j \leqq 4)$$

$$x_7 = \mathcal{P}_3^1 x_3, \qquad x_{15} = \mathcal{P}_3^1 x_{11};$$

Enfin, $\psi^*(H^8(\boldsymbol{F}_4, Z))$ est $\neq 0$ donc (20.7) d'ordre 3, d'où le fait que pour un certain multiple x'_8 de x_8 on a

$$\delta^*(x_7) = x'_8, \qquad \delta^*(x_7 \cdot x'_8) = x'^2_8$$

(δ^*, homomorphisme de Bockstein), ce qui termine la démonstration de 19.3.

Remarque. Nous venons de voir que $H^8(\boldsymbol{F}_4, Z) \neq 0$, ce qui, vu (20.7) donne:

$$H^8(\boldsymbol{F}_4, Z) = Z_3. \tag{21.1}$$

22. Cohomologie mod. 2 de $\boldsymbol{F}_4$. $H^*(\boldsymbol{Spin}(9), Z_2)$ possède un système simple de générateurs universellement transgressifs de degrés 3, 5, 6, 7, 15, (15.2), et d'autre part $\psi^*(H^8(\boldsymbol{W}, Z_2)) = 0$ d'après (20.5); on pourrait en déduire comme précédemment la structure de l'algèbre spectrale sur Z_2 de la fibration $(\boldsymbol{F}_4, \boldsymbol{W}, \boldsymbol{Spin}(9), \psi)$ et le fait que $H^*(\boldsymbol{F}_4, Z_2)$ a un système simple de générateurs de degrés 3, 5, 6, 15, 23. Cependant, pour obtenir quelques précisions supplémentaires sur les i-carrés et la transgression, nous retrouverons et compléterons ce résultat par une méthode légèrement différente, qui détermine aussi la cohomologie mod. 2 de $\boldsymbol{Spin}(9)$, mais utilise celle de $\boldsymbol{G}_2$. Nous choisissons une fois pour toutes une suite de sous-groupes emboîtés:

$$\boldsymbol{F}_4 \supset \boldsymbol{Spin}(9) \supset \boldsymbol{Spin}(7) \supset \boldsymbol{G}_2$$

tels que $\boldsymbol{Spin}(9)/\boldsymbol{Spin}(7) = \boldsymbol{S}_{15}$ et $\boldsymbol{Spin}(7)/\boldsymbol{G}_2 = \boldsymbol{S}_7$, (*voir* [1]), et soient λ, μ les projections naturelles

$$\boldsymbol{F}_4 \xrightarrow{\lambda} \boldsymbol{F}_4/\boldsymbol{G}_2 \xrightarrow{\mu} \boldsymbol{F}_4/\boldsymbol{Spin}(9) = \boldsymbol{W}.$$

LEMME 22.1. $H^*(\boldsymbol{Spin}(9)/\boldsymbol{G}_2, Z) \cong H^*(\boldsymbol{S}_7 \times \boldsymbol{S}_{15}, Z) \cong \wedge(u_7, u_{15})$, $(Du_i = i)$.

En effet, l'algèbre spectrale sur Z de la fibration

$$(\boldsymbol{Spin}(9)/\boldsymbol{G}_2, \boldsymbol{Spin}(9)/\boldsymbol{Spin}(7), \boldsymbol{Spin}(7)/\boldsymbol{G}_2) = (\boldsymbol{Spin}(9)/\boldsymbol{G}_2, \boldsymbol{S}_{15}, \boldsymbol{S}_7)$$

est visiblement triviale.

LEMME 22.2. *$H^i(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) = 0$ pour $0 < i < 7$, $i = 9$; λ^* est un isomorphisme de $H^8(\boldsymbol{F}_4/\boldsymbol{G}_2, Z)$ sur $H^8(\boldsymbol{F}_4, Z)$; en particulier, vu* (21.1), $H^8(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) \cong Z_3$.

La première assertion résulte du fait que dans l'algèbre spectrale sur Z de la fibration $(\boldsymbol{F}_4/\boldsymbol{G}_2, \boldsymbol{W}, \boldsymbol{Spin}(9)/\boldsymbol{G}_2, \mu)$ le terme E_2 n'a aucun élément non nul de degré total i, pour $0 < i < 7$, $i = 9$.

On en déduit, à l'aide de (17.2), que dans l'algèbre spectrale sur Z de $(\boldsymbol{F}_4, \boldsymbol{F}_4/\boldsymbol{G}_2, \boldsymbol{G}_2, \lambda)$:

$$E_2^{p,q} = 0, \quad (p + q = 7, q > 0); \qquad E_2^{8,0} = {}^8E_2,$$

ce qui implique

$$E_2^{8,0} \cong E_\infty^{8,0} = {}^8E_\infty$$

donc que λ^* est un isomorphisme sur en dimension 8.

LEMME 22.3. *Soit $p \neq 3$. Alors $H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z_p) = H^*(\boldsymbol{S}_{15} \times \boldsymbol{S}_{23}, Z_p)$; en particulier, $\boldsymbol{F}_4/\boldsymbol{G}_2$ n'a pas de p-torsion.*

Dans l'algèbre spectrale mod. p de $(\boldsymbol{F}_4/\boldsymbol{G}_2, \boldsymbol{W}, \boldsymbol{Spin}(9)/\boldsymbol{G}_2, \mu)$, on a

$$E_2 \cong H^*(\boldsymbol{W}, Z_p) \otimes H^*(\boldsymbol{Spin}(9)/\boldsymbol{G}_2, Z_p) \cong Z_p[x]/(x^3) \otimes \wedge(u_7, u_{15})$$

$(Dx = 8)$, d'où

$$E_2 \cong E_8\,; \qquad d_8\kappa_8{}^2(1 \otimes u_{15}) = 0\,;$$

mais le lemme précédent montre que

$$H^8(\boldsymbol{F}_4/\boldsymbol{G}_2, Z_p) \cong H^8(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) \otimes Z_p = 0$$

ce qui n'est possible que si

$$d_8\kappa_8{}^2(1 \otimes u_7) = \kappa_8{}^2(x \otimes 1)\,;$$

on en déduit immédiatement que

$$E_\infty \cong E_9 \cong \wedge(x^2 \otimes u_7, u_{15})$$

et le lemme.

PROPOSITION 22.4. *Le groupe $\boldsymbol{G}_2$ est totalement non homologue à zéro* mod. 2 *dans $\boldsymbol{F}_4$; $H^*(\boldsymbol{F}_4, Z_2)$ est isomorphe à $H^*(\boldsymbol{G}_2 \times \boldsymbol{S}_{15} \times \boldsymbol{S}_{23}, Z_2)$ pour la structure additive et le cup-produit, et, au moins pour $D \leqq 22$, au sens des i-carrés.*

Nous désignerons par i l'inclusion de $\boldsymbol{G}_2$ dans $\boldsymbol{F}_4$ et par y_3, y_5, y_6 les générateurs universellement transgressifs de $H^*(\boldsymbol{G}_2, Z_2)$.

$H^4(\boldsymbol{F}_4/\boldsymbol{G}_2, Z_2)$ étant nul d'après (22.2), l'image par transgression dans $\boldsymbol{F}_4$ de y_3 est nulle, autrement dit y_3 fait partie de l'image de i^*, et il ne sera de même pour

$$y_5 = Sq^2y_3, \qquad y_6 = y_3 \cdot y_3.$$

Ainsi, $\boldsymbol{G}_2$ est totalement non homologue à zéro mod. 2 dans $\boldsymbol{F}_4$, et dans l'algèbre spectrale de $(\boldsymbol{F}_4, \boldsymbol{F}_4/\boldsymbol{G}_2, \boldsymbol{G}_2, \lambda)$, on a, vu (22.3) :

$$E_\infty \cong E_2 \cong H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z_2) \otimes H^*(\boldsymbol{G}_2, Z_2) \cong \wedge(y_{15}, y_{23}) \otimes \Delta(y_3, y_5, y_6)$$

$(Dy_i = i)$, et ([3], Prop. 8.1), $H^*(\boldsymbol{F}_4, Z_2)$ possède un système simple de générateurs $x_3, x_5, x_6, x_{15}, x_{23}$ tels que

$$(22.5)\qquad \begin{aligned} i^*(x_i) &= y_i, && (i = 3, 5, 6),\\ \lambda^*(y_i) &= x_i, && (i = 15, 23). \end{aligned}$$

En fait, pour des raisons de degrés évidentes, les éléments x_3, x_5, x_6 sont *univoquement* déterminés par 22.5, et engendrent une sous-algèbre isomorphe à $H^*(\boldsymbol{G}_2, Z_2)$, ce qui, compte tenu de (7.4) et de (17.3), donne

$$H^*(\boldsymbol{F}_4, Z_2) \cong E_\infty = H^*(\boldsymbol{G}_2 \times \boldsymbol{S}_{15} \times \boldsymbol{S}_{23}, Z_2)$$

$Sq^2 x_3 = x_5$, $Sq^1 x_5 = x_6$, $Sq^3 x_3 = x_6$, $Sq^i x_j = 0$ sinon, $(j \leqq 6)$, et, joint aux égalités

$$Sq^i x_{15} = \lambda^*(Sq^i y_{15}) = 0 \qquad (0 < i \leqq 7),$$

$$Sq^i x_{23} = \lambda^*(Sq^i y_{23}) = 0 \qquad (i > 0),$$

démontre (22.4), (19.5), et (19.2(c)), mis à part cependant le fait que les x_i sont universellement transgressifs, qui fait l'objet de la Proposition 22.7.

LEMME 22.6. *Les générateurs y_{15} et y_{23} de $H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z_2)$ sont transgressifs dans la fibration $(B_{\boldsymbol{G}_2}, B_{\boldsymbol{F}_4}, \boldsymbol{F}_4/\boldsymbol{G}_2, \rho(\boldsymbol{G}_2, \boldsymbol{F}_4))$.*

Dans l'algèbre spectrale mod. 2 de cette fibration, on a évidemment

$$E_r^{p,q} = 0, \qquad (q \neq 0, 15, 23, r \geqq 2),$$

y_{15} est donc transgressif et de plus

$$E_2^{p,q} \cong E_\infty^{p,q}, \qquad (p + q < 15),$$

en particulier ${\rho^*}_2(\boldsymbol{G}_2, \boldsymbol{F}_4)$ est un isomorphisme sur pour $D < 15$; mais d'après 17.3 et la Prop. 19.1 de [3] :

$$H^*(B_{\boldsymbol{G}_2}, Z_2) \cong Z_2[v_4, v_6, v_7] \qquad (Dv_i = i),$$

d'où

$$E_r^{9,0} = E_r^{9,15} = 0 ;$$

ainsi, nous voyons que

$$d_r(E_r^{0,23}) \subset E_r^{r,24-r} = 0, \qquad (r \geqq 2),$$

et y_{23} est transgressif.

PROPOSITION 22.7. *Les générateurs x_i $(i = 3, 5, 6, 15, 23)$, de $H^*(\boldsymbol{F}_4, Z_2)$ définis par (22.5) sont universellement transgressifs.*

Cela est clair pour x_3, qui a le plus petit degré > 0, donc pour $x_5 = Sq^2 x_3$ et $x_6 = x_3 \cdot x_3$. Considérons maintenant le diagramme

$$\begin{array}{ccc} E_{F_4} & \xrightarrow{\nu} & B_{G_2} \\ \downarrow & & \downarrow \rho(G_2, F_4) \\ B_{F_4} & \xrightarrow{\text{id.}} & B_{F_4} \end{array}$$

où ν est la projection naturelle:

$$E_{F_4} = E_{G_2} \xrightarrow{\nu} B_{G_2} = E_{G_2}/G_2.$$

la restriction de ν à une fibre G_2 est λ et comme y_{15}, y_{23} sont transgressifs dans la fibration de droite, les éléments $x_i = \lambda^*(y_i)$, $(i = 15, 23)$ seront transgressifs dans la fibration de gauche, c'est à dire universellement transgressifs.

23. Cohomologie entière de F_4.

PROPOSITION 23.1. *$H^*(F_4/G_2, Z)$ est additivement et multiplicativement isomorphe au produit $U \otimes H^*(S_{15}, Z)$, où U est une algèbre unitaire graduée définie par $U^0 \cong U^{23} \cong Z$, $U^8 \cong U^8 \cdot U^8 \cong U^{16} \cong Z_3$, $U^i = 0$ sinon.*

On a vu (22.2) que $\lambda^*: H^8(F_4/G_2, Z) \to H^8(F_4, Z)$ est un isomorphisme sur, et comme $H^8(F_4, Z_3) \cong H^8(F_4, Z) \otimes Z_3$ est engendré par un élément de carré non nul, il en sera de même pour $H^8(F_4, Z)$ et $H^8(F_4/G_2, Z)$; cela va nous permettre de construire l'algèbre spectrale sur Z de la fibration $(F_4/G_2, W, Spin(9)/G_2, \mu)$. On a, compte tenu de (22.1),

$$E_2 \cong H^*(W, Z) \otimes H^*(Spin(9)/G_2, Z) \cong Z[x]/(x^3) \otimes \wedge(u_7, u_{15}) = E_8$$

$$d_8\kappa_8{}^2(1 \otimes u_7) = \kappa_8{}^2(3x \otimes 1),$$

$$d_8\kappa_8{}^2(1 \otimes u_{15}) \in E_8{}^{8,8} = 0,$$

les d_8-cocycles forment donc la sous-algèbre

$$x^2 \otimes u_7 \otimes \wedge(u_{15}) + Z[x]/(x^3) \otimes \wedge(u_{15})$$

et les d_8-cobords forment

$$(3x) \otimes \wedge(u_{15}),$$

ce qui donne

$$E_9 \cong H^*(W, Z)/(3x) \otimes \wedge(u_{15}) + (x^2 \otimes u_7) \otimes \wedge(u_{15}) = U \otimes \wedge(u_{15}).$$

Comme $E_\infty{}^{16,0} \neq 0$, il faut que $d_{15}\kappa_{15}{}^2(1 \otimes u_{15})$ soit nul, ce qui entraîne $E_9 \cong E_\infty$. Ainsi, E_∞ est formé d'un groupe Z en dimensions 0, 15, 23, 38 et d'un groupe Z_3 en dimensions 8, 16, 23, 31, ce qui montre déjà

$$H^i(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) \cong {}^iE_\infty \quad \text{pour } i \neq 23$$

$$H^{23}(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) \cong Z \text{ ou } Z + Z_3;$$

mais par dualité de Poincaré, puisque dim. $\boldsymbol{F}_4/\boldsymbol{G}_2 = 38$,

$$\text{Tors. } H^{23}(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) = \text{Tors. } H^{16}(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) = \text{Tors. } {}^{16}E_\infty = Z_3$$

ce qui prouve que $H^{23}(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) \cong Z + Z_3 \cong {}^{23}E_\infty$, donc que $H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z)$ est additivement isomorphe à E_∞; que cet isomorphisme soit multiplicatif résulte alors directement des faits suivants: (a) E_∞ est l'algèbre graduée associée à $H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z)$ convenablement filtrée, (b) μ^* est un isomorphisme de $H^*(\boldsymbol{W}, Z)/(3x)$ dans $H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z)$, (c) des représentants quelconques dans $H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z)$ de u_{15} et $x^2 \otimes u_7$ sont forcément de carré nul.

PROPOSITION 23. 2. *Dans les notations de* (22. 1), $H^*(\boldsymbol{F}_4, Z)$ *est additivement et multiplicativement isomorphe à* $U \otimes H^*(\boldsymbol{G}_2 \times \boldsymbol{S}_{15}, Z)$.

Tors. $H^*(\boldsymbol{G}_2, Z)$ ne contenant que des éléments d'ordre 2, le résultat précédent montre que

$$\text{Tor.}(H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z), H^*(\boldsymbol{G}_2, Z)) = 0;$$

dans l'algèbre spectrale sur Z de la fibration $(\boldsymbol{F}_4, \boldsymbol{F}_4/\boldsymbol{G}_2, \boldsymbol{G}_2, \lambda)$ on a donc

$$(23. 3) \qquad E_2 \cong H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) \otimes H^*(\boldsymbol{G}_2, Z) \cong U \otimes \wedge(u_{15}) \otimes H^*(\boldsymbol{G}_2, Z).$$

Le plus petit degré-base > 0 est 8, les différentielles d_r $(2 \leqq r \leqq 7)$, sont donc nulles; par ailleurs $H^*(\boldsymbol{G}_2, Z)$ est engendré par deux éléments y_3, y_{11} d'après (17. 2); les différentielles d_r $(r \geqq 8)$, sont évidemment nulles sur ${\kappa_r}^2(1 \otimes y_3)$, de plus elles annulent aussi ${\kappa_r}^2(1 \otimes y_{11})$ car ${}^{12}E_r = 0$; ainsi finalement $d_r = 0$ pour tout $r \geqq 2$, E_∞ est isomorphe à E_2 et ses coefficients de torsion sont égaux à 2, 3, ou 6; en explicitant (23. 3) on obtient:

$${}^{12}E_\infty = {}^{17}E_\infty = {E_\infty}^{16,6} = 0$$

$${}^{22}E_\infty = {E_\infty}^{8,14} = Z_3,$$

par conséquent:

$$(23. 4) \qquad H^{12}(\boldsymbol{F}_4, Z) = H^{17}(\boldsymbol{F}_4, Z) = 0; \qquad H^{22}(\boldsymbol{F}_4, Z) = Z_3.$$

Soit $x'_j \in H^j(\boldsymbol{F}_4, Z)$ tel que

$$i^*(x'_j) = y_j, \qquad (j = 3, 11, i \text{ inclusion de } \boldsymbol{G}_2 \text{ dans } \boldsymbol{F}_4),$$

y_j étant le générateur de $H^*(\boldsymbol{G}_2, Z)$ mentionné dans (17. 2). Les égalités (23. 4) montrent que

$${x'_3}^4 = {x'_3}^2 \cdot x'_{11} = 0$$

et aussi que $x'_{11}{}^2 = 0$ car si cet élément était non nul, il devrait être d'ordre deux; x'_3 et x'_{11} engendrent donc une sous-algèbre isomorphe à $H^*(\boldsymbol{G}_2, Z)$, et l'isomorphie additive et multiplicative de $H^*(\boldsymbol{F}_4, Z)$ avec

$$E_\infty \cong E_2 \cong U \otimes H^*(\boldsymbol{G}_2 \times \boldsymbol{S}_{15}, Z)$$

résulte de (7. 4).

Remarque. Nous avons vu que dans l'algèbre spectrale sur Z de la fibration $(\boldsymbol{F}_4, \boldsymbol{F}_4/\boldsymbol{G}_2, \boldsymbol{G}_2, \lambda)$

$$E_\infty \cong E_2 \cong H^*(\boldsymbol{F}_4/\boldsymbol{G}_2, Z) \otimes H^*(\boldsymbol{G}_2, Z)$$

par conséquent, $\boldsymbol{G}_2$ est totalement non homologue à zéro dans $\boldsymbol{F}_4$ relativement à la cohomologie entière.

THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHIE.

[1] A. Borel, "Le plan projectif des octaves et les sphères comme espaces homogènes," *Comptes rendus de l'Académie des Sciences*, Paris, t. 230 (1950), pp. 1378-1380.

[2] ———, "Sur la cohomologie des variétés de Stiefel et de certains groupes de Lie," *ibid.*, t. 232 (1950), pp. 1628-1630.

[3] ———, "Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts," *Annals of Mathematics*, t. 57 (1953), pp. 115-207.

[4] ———, "La cohomologie mod. 2 de certains espaces homogènes," *Commentarii Mathematici Helvetici*, t. 27 (1953), pp. 165-197.

[5] ——— et J.-P. Serre, "Sur certains sous-groupes des groupes de Lie compacts," *ibid.*, t. 27 (1953), pp. 128-139.

[6] ——— et J.-P. Serre, "Groupes de Lie et puissances réduites de Steenrod," *American Journal of Mathematics*, t. 75 (1953), pp. 409-448.

[7] C. Chevalley, "Determination of the Betti numbers of the exceptional Lie groups," *Proceedings of the International Congress of Mathematicians*, 1950, t. II, pp. 21-24.

[8] C. Ehresmann, "Sur la variété des génératrices planes d'une quadrique réelle et sur la topologie du groupe orthogonal," *Comptes rendus de l'Académie des Sciences*, Paris, t. 208 (1939), pp. 321-323.

[9] ———, "Sur la topologie des groupes simples clos," *ibid.*, pp. 1263-1265.

[10] H. Hopf, "Ueber die Topologie der Gruppen-Mannigfaltigkeiten und ihrer Verallgemeinerungen," *Annals of Mathematics*, t. 42 (1941), pp. 22-52.

[11] J. Leray, "Sur la forme des espaces topologiques et sur les points fixes des représentations," *Journal de Mathématiques Pures et Appliquées*, t. 54 (1945), pp. 95-167.

[12] ———, "Espaces où opère un groupe de Lie compact connexe," *Comptes rendus de l'Académie des Sciences, Paris*, t. 228 (1949), pp. 1545-1547.

[13] ———, "Applications continue commutant avec les éléments d'un groupe de Lie," *ibid.*, pp. 1784-1786.

[14] ———, "Sur l'homologie des groupes de Lie, des espaces homogènes et des espaces fibrés principaux," *Colloque de Topologie (espaces fibrés)*, Bruxelles, 1950, pp. 101-115.

[15] C. E. Miller, "The topology of rotation groups," *Annals of Mathematics*, t. 57 (1953), pp. 90-114.

[16] L. Pontrjagin, "Homologies in compact Lie groups," *Matematicheski Sbornik*, t. 6 (1939), pp. 389-422.

[17] H. Samelson, "Beiträge zur Topologie der Gruppen-Mannigfaltigkeiten," *Annals of Mathematics*, t. 42 (1941), pp. 1091-1137.

[18] Yen Chih-Tah, "Sur les polynômes de Poincaré des groupes de Lie exceptionnels," *Comptes rendus de l'Académie des Sciences*, Paris, t. 228 (1949), pp. 628-630.

ON A PROBLEM OF PARTITIONS II.*

By ALFRED BRAUER and B. M. SEELBINDER.

A number of years ago, the first of the authors studied together with I. Schur the following problem of Frobenius: Given k relatively prime positive integers $a_1, a_2, \cdots, a_k$, find bounds $F(a_1, a_2, \cdots, a_k)$ such that the linear Diophantine equation

$$n = a_1x_1 + a_2x_2 + \cdots + a_kx_k \tag{1}$$

always has solutions in positive integers $x_1, x_2, \cdots, x_k$ for $n > F(a_1, a_2, \cdots, a_k)$.

In a paper appearing in vol. 64 (1942), pp. 299-312 of this JOURNAL, the following results were obtained: Let $d_1 = a_1$, $d_2 = (a_1, a_2), \cdots, d_k = (a_1, a_2, \cdots, a_k) = 1$ be the greatest common divisors. Then

$$T = T(a_1, a_2, \cdots, a_k) = a_2d_1/d_2 + a_3d_2/d_3 + \cdots, + a_kd_{k-1}/d_k$$

is such a bound $F(a_1, a_2, \cdots, a_k)$. For $k = 2$ this bound becomes $T = a_1a_2$ and it is the best possible bound. But T is not, in general, the best bound. For instance, if $a_1 = m$, $a_2 = m + 1$, $a_3 = m + 2$ are three consecutive integers and m is even, then $T(m, m + 2, m + 1) = m(m + 2)/2 + 2(m + 1) = (m^2 + 6m + 4)/2$ is the best possible bound, while

$$\min \{T(m, m + 1, m + 2), T(m, m + 2, m + 1), T(m + 1, m + 2, m)\} = m^2 + 2m + 2$$

does not give the best bound for odd m. Here $(m^2 + 5m + 4)/2$ is the best bound, since for the case of k consecutive integers

$$U(m, m + 1, m + 2, \cdots, m + k - 1) = [(m + k^2 - 3)/(k - 1)]m + (k^2 - k - 2)/2$$

was shown to be the best bound.[1] Moreover, it was proved that T is the best bound if the integers a_κ/d_κ are representable in the form

$$a_\kappa/d_\kappa = (\sum_{\nu=1}^{\kappa-1} a_\nu y_{\kappa\nu})/d_{\kappa-1} \tag{2}$$

* Received August 14, 1953.

[1] Loc. cit. Theorem 7, p. 310; on p. 301, l. 6, the factor m is missing in the formula for U. On p. 306, l. 20, the summation should extend to κ instead of k.

with $y_{\kappa\nu} \geqq 0$ for $\kappa = 3, 4, \cdots, k$. For $k = 3$ the condition is not only sufficient but also necessary in order that T be the best bound. It was stated (loc. cit., p. 301) that the condition (2) is not necessarily for $k > 3$, but this is not correct.

It will be proved in this paper that the condition (2) is necessary for every k in order that T be the best bound. Moreover, if (2) is not satisfied, then $T^* = T - \min(a_1, a_2, \cdots, a_k) = F(a_1, a_2, \cdots, a_k)$ is a smaller bound for the solvability of (1) in positive integers. First we prove the following:

LEMMA. *Let $a_1, a_2, \cdots, a_k$ be relatively prime positive integers. Then every integer m is representable in at least one of the forms*

$$(3) \quad m = a_1x_1 + a_2x_2 + \cdots + a_kx_k \qquad (x_\kappa > 0 \text{ for } \kappa = 1, 2, \cdots, k),$$

$$(4) \quad m = T - a_1x_1 - a_2x_2 - \cdots - a_kx_k \quad (x_\kappa \geqq 0 \text{ for } \kappa = 1, 2, \cdots, k).$$

Proof. Since $(a_k, d_{k-1}) = 1$, the congruence $a_kz_k \equiv m \pmod{d_{k-1}}$ has a solution such that $0 < z_k \leqq d_{k-1}$. We set $m = a_kz_k + d_{k-1}g_{k-1}$. Next consider the congruence

$$(5) \qquad a_{k-1}z_{k-1} \equiv d_{k-1}g_{k-1} \pmod{d_{k-2}}.$$

Because $d_{k-2} = (a_1, a_2, \cdots, a_{k-2})$ and $d_{k-1} = (a_1, a_2, \cdots, a_{k-1}) = (d_{k-2}, a_{k-1})$, we can divide (5) by d_{k-1} and obtain

$$(6) \qquad a_{k-1}z_{k-1}/d_{k-1} \equiv g_{k-1} \pmod{d_{k-2}/d_{k-1}}.$$

Now $(a_{k-1}/d_{k-1}, d_{k-2}/d_{k-1}) = 1$, hence (6) and (5) have a solution z_{k-1} such that $0 < z_{k-1} \leqq d_{k-2}/d_{k-1}$. Hence we may set $m = a_kz_k + a_{k-1}z_{k-1} + d_{k-2}g_{k-2}$. We continue in this way and obtain by induction

$$(7) \qquad m = a_kz_k + a_{k-1}z_{k-1} + \cdots + a_2z_2 + d_1g_1,$$

where

$$(8) \qquad 0 < z_{k-i} \leqq d_{k-i-1}/d_{k-i}, \qquad (i = 0, 1, \cdots, k-2).$$

If $g_1 > 0$, then it follows from (7) that m is representable in form (3), since $d_1 = a_1$. Assume now that $g_1 \leqq 0$. Let $g_1 = -h$. Adding and subtracting T in equation (7) we obtain

$$m = T - a_2a_1/d_2 - a_3d_2/d_3 - \cdots - a_kd_{k-1}/d_k + a_kz_k + a_{k-1}z_{k-1} + \cdots + a_2z_2 - d_1h$$

or, replacing d_1 by a_1,

$$m = T - a_1 h - a_2(a_1/d_2 - z_2) - a_3(d_2/d_3 - z_3) - \cdots - a_k(d_{k-1}/d_k - z_k).$$

Since all the parentheses are non-negative by (8), it follows that m is representable in the form (4).

THEOREM 1. *If at least one of the integers a_κ/d_κ $(\kappa = 3, 4, \cdots, k)$ is not representable in the form*

$$a_\kappa/d_\kappa = \sum_{\nu=1}^{\kappa-1} a_\nu y_{\kappa\nu}/d_{\kappa-1} \text{ with } y_{\kappa\nu} \geqq 0, \tag{9}$$

then $T(a_1, a_2, \cdots, a_k)$ is not the smallest bound.

Proof. For $k = 3$ this was proved *loc. cit.* Let us assume that the theorem is true for $k - 1$. Let κ be the smallest integer for which (9) does not hold. We have to consider two cases.

The case $\kappa < k$. Set

$$\begin{aligned} T_\kappa &= T(a_1/d_\kappa,\ a_2/d_\kappa, \cdots,\ a_\kappa/d_\kappa) \\ &= (a_2 d_1/d_2 + a_3 d_2/d_3 + \cdots + a_\kappa d_{\kappa-1}/d_\kappa)/d_\kappa. \end{aligned} \tag{10}$$

It follows from the inductive assumption that T_κ is not the smallest bound. Hence the equation

$$T_\kappa = (a_1 x_1 + a_2 x_2 + \cdots + a_\kappa x_\kappa)/d_\kappa \tag{11}$$

has a positive solution $x_1, x_2, \cdots, x_\kappa$. On the other hand, by (10),

$$\begin{aligned} T &= T(a_1, a_2, \cdots, a_k) \\ &= d_\kappa T_\kappa + a_{\kappa+1} d_\kappa/d_{\kappa+1} + a_{\kappa+2} d_{\kappa+1}/d_{\kappa+2} + \cdots + a_k d_{k-1}/d_k. \end{aligned} \tag{12}$$

Hence

$$\begin{aligned} T = a_1 x_1 + a_2 x_2 + \cdots + a_\kappa x_\kappa \\ + a_{\kappa+1} d_\kappa/d_{\kappa+1} + a_{\kappa+2} d_{\kappa+1}/d_{\kappa+2} + \cdots + a_k d_{k-1}/d_k, \end{aligned}$$

by (11). Since $d_\kappa/d_{\kappa+1}$, $d_{\kappa+1}/d_{\kappa+2}, \cdots, d_{k-1}/d_k$ are positive integers, T is not the best bound.

The case $\kappa = k$. Since a_k cannot be represented in the form (9), the number $T_{k-1} - a_k$ cannot be written in the form

$$T_{k-1} - \sum_{\nu=1}^{k-1} a_\nu y_{k\nu}/d_{k-1} \qquad (y_{k\nu} \geqq 0).$$

Therefore, by the Lemma, $T_{k-1} - a_k$ can be represented in the form $T_{k-1} - a_k = (a_1x_1 + a_2x_2 + \cdots + a_{k-1}x_{k-1})/d_{k-1}$, where $x_1, x_2, \cdots, x_{k-1}$ are positive integers. Hence, by (12),

$$T = d_{k-1}T_{k-1} + a_k d_{k-1}/d_k = a_1x_1 + a_2x_2 + \cdots + a_{k-1}x_{k-1} + 2a_k d_{k-1}.$$

Choosing $x_k = 2d_{k-1}$ we obtain the representation for T. Hence T is not the best bound.

THEOREM 2. *If T is not the best bound, then the best bound is less than or equal to $T - \min(a_1, a_2, \cdots, a_k) = T^*$.*

Proof. We can formulate the Lemma in the following way: If for an integer m the number $T - m$ cannot be represented in the form

$$T - m = a_1x_1 + a_2x_2 + \cdots + a_kx_k \qquad (x_\kappa \geqq 0), \tag{13}$$

then m can be represented in the form (3). Let r be any of the numbers $1, 2, \cdots, \min(a_1, a_2, \cdots, a_k) - 1$. We set $T - r = R$. Since

$$r < \min(a_1, a_2, \cdots, a_k),$$

$r = T - R$ cannot be represented in the form (13) and it follows that R can be represented in the form (3). Hence all of the numbers greater than T^* can be represented in the form (3).

UNIVERSITY OF NORTH CAROLINA.

FOOTNOTE TO A RECENT PAPER.*

By André Weil.

In a recent paper, L. Carlitz [1] has given formulas for the number of solutions of special pairs of equations $F = a$, $G = b$, where F, G are quadratic forms over a field k with q elements, q being odd. More complete results may be obtained more briefly by the same method. Observe first that, if N is the number of solutions of $F = a$, $G = b$, and if N', N'' are the numbers of solutions of $F = G = 0$ and of $F - at^2 = G - bt^2 = 0$, respectively, where t is a variable not occurring in F, G, then $N = (q-1)^{-1}N'' - N'$; so it is enough to consider the homogeneous case $F = G = 0$, where F, G are quadratic forms in any number of variables $x_0, x_1, \cdots, x_n$, which may of course be assumed to be linearly independent. By $\det(F)$, if $F = \sum a_{ij}x_ix_j$, we understand the determinant of the coefficients a_{ij}. We put

$$\phi(u, v) = \det(uF + vG);$$

this is a form of degree $n+1$ in u, v, which may be identically 0.

Let ψ be any non-trivial character of the additive group of the elements in k; let χ be the character of order 2 of the multiplicative group of non-zero elements in k, i. e. $\chi(x) = 1$ or -1 according as x is or is not a square in k, and $\chi(0) = 0$. Then the Gaussian sum $g = \sum \chi(x)\psi(x)$ satisfies $g^2 = \chi(-1)q$. Carlitz's method depends upon the following well-known and elementary fact: if the quadratic form $H(x_0, \cdots, x_n)$ is equivalent over k to the form $\sum_{i=0}^{m} c_i y_i^2$, where the c_i are $\neq 0$ and $0 \leqq m \leqq n$, then

$$\sum_x \psi[H(x_0, \cdots, x_n)] = \chi(c_0c_1 \cdots c_m)g^{m+1}q^{n-m},$$

so that in particular, if $\det(H) \neq 0$ and therefore $m = n$, the left-hand side has the value $\chi[\det(H)]g^{n+1}$.

Now, calling N the number of solutions of $F = G = 0$, we observe with Carlitz that we have

$$N = q^{-2} \sum_{x,u,v} \psi(uF)\psi(vG) = q^{-2} \sum_{u,v} \left(\sum_x \psi(uF + vG)\right),$$

* Received November 27, 1953.

and evaluate the last sum by the above formula. For $u = v = 0$, we get q^{n+1}. For $\phi(u, v) \neq 0$, we get $\chi[\phi(u, v)]g^{n+1}$. Let now $b_\rho u - a_\rho v$, for $1 \leqq \rho \leqq r$, be all the essentially distinct linear forms with coefficients in k which divide $\phi(u, v)$; two linear forms which differ only by a constant factor are not to be reckoned as essentially different. If ϕ is identically 0 we have $r = q + 1$, otherwise $r \leqq n + 1$. Let $m_\rho + 1$ be the rank of $H_\rho = a_\rho F + b_\rho G$; let H_ρ be equivalent over k to $\sum_i c_{\rho i} y_i^2$, where $0 \leqq i \leqq m_\rho$, and put $C_\rho = c_{\rho 0} c_{\rho 1} \cdots c_{\rho m_\rho}$. Then we get

$$N = q^{n-1} + q^{-2} g^{n+1} S + \sum_\rho q^{n-m_\rho-2} g^{m_\rho+1} T_\rho$$

where S and the T_ρ are defined by

$$S = \sum_{\phi(u,v)\neq 0} \chi[\phi(u, v)], \qquad T_\rho = \sum_{t\neq 0} \chi(C_\rho t^{m_\rho+1}).$$

Clearly the latter sum is 0 if m_ρ is even, and has the value $(q-1)\chi(C_\rho)$ if m_ρ is odd. In the former sum, the restriction $\phi(u, v) \neq 0$ may be removed, as $\chi(0) = 0$; replacing u, v in it by ut, vt changes S into $\chi(t^{n+1})S$, so that S is 0 if n is even. If n is odd, it is clear that we have $S = \bar{N} - q^2$, where $\bar{N}$ is the number of solutions of $t^2 = \phi(u, v)$ and is related in an obvious manner to the number of points on the hyperelliptic curve $y^2 = \phi(x, 1)$. Specializing this to the case considered by Carlitz, one gets all his results immediately.

For a non-singular variety V in a projective space over a field k with q elements, I have defined ([2], p. 507) a zeta-function $Z(U)$ by the formula

$$D \log Z(U) = \sum_1^\infty N_\nu U^{\nu-1}, \qquad (D = d/dU),$$

where N_ν is the number of points on V with coordinates in the extension k_ν of degree ν of k. Now, for the variety defined by the pair of quadratic equations $F = G = 0$ to be non-singular, it is necessary and sufficient that $\phi(u, v)$ should not be identically 0 and should have no multiple linear factors in any extension of k (i. e., that its discriminant should be $\neq 0$). If that is so, then, with the same notations as above, m_ρ cannot have any other value than $n - 1$. If therefore n is odd, the T_ρ are all 0. The curve $y^2 = \phi(x, 1)$ has here the genus $(n-1)/2$, and its zeta-function is known to be of the form

$$\bar{Z}(U) = (1 - U)^{-1}(1 - qU)^{-1} \prod_{\nu=1}^{n-1} (1 - \alpha_\nu U)$$

where the α_ν are of absolute value $q^{1/2}$. Then the above formulas give, for the zeta-function $Z(U)$ of the variety $F = G = 0$, the value

$$Z(U) = \prod_{h=0}^{n-2} (1 - q^h U)^{-1} \prod_{\nu=1}^{n-1} (1 - q^{(n-3)/2} \chi(-1)^{(n+1)/2} \alpha_\nu U);$$

this seems to suggest some relationship between the intrinsic properties of the variety $F = G = 0$ and those of the curve $y^2 = \phi(x, 1)$. On the other hand, one finds, when n is even:

$$Z(U) = \prod_{h=0}^{n-2} (1 - q^h U)^{-1} \prod_{\lambda} (1 - q^{d_\lambda (n-2)/2} \chi(-1)^{d_\lambda n/2} \epsilon_\lambda U^{d_\lambda})^{-1},$$

where the d_λ and ϵ_λ are as follows. The d_λ are the degrees of all the irreducible factors $\phi_\lambda(u, v)$ into which $\phi(u, v)$ splits up over the field k; and, if k_λ is the extension of k of degree λ, and $b_\lambda u - a_\lambda v$ is a linear factor of $\phi_\lambda(u, v)$ in k_λ, ϵ_λ is 1 or -1 according as $a_\lambda F + b_\lambda G$ is or is not equivalent over k_λ to a sum of n squares.

The same method still yields interesting results if one applies it to any number of homogeneous quadratic equations $F_1 = \cdots = F_r = 0$, provided the quadratic forms $F_1, \cdots, F_r$ contain only the squares of the variables (this is the assumption made by Carlitz in the case $r = 2$, unnecessarily as we have shown). It gives a relation between the number of solutions of that system and the numbers of solutions of certain equations of the type $t^2 = P(u_1, \cdots, u_s)$ where the right-hand side is a product of an even number of linear forms. In the case of a non-singular variety defined by three quadratic equations of the above form, one gets in this manner a fairly simple formula for the zeta-function.

The above results provide additional evidence concerning the conjectures on the zeta-functions of non-singular varieties formulated in my article [2]; for a comparison with the topological properties of varieties defined by pairs of quadratic equations $F = G = 0$ over complex numbers, results of Hirzebruch (still unpublished) are now available, which include the calculation of the Betti numbers of all non-singular complete intersections of hypersurfaces in a projective space. Since somewhat misleading statements have crept into the recent literature on this subject, it seems worthwhile to emphasize that *all the evidence on hand* (including some which is still unpublished) *fully confirms those conjectures,* not only as to the rationality of the zeta-functions but also as to the existence of the functional equation and the relationship with Betti numbers in the classical case.

Naturally, one may also attach a function $Z(U)$, defined by the same

formula as above, to varieties with singular points, to varieties in affine spaces, etc., and it seems very likely that all functions defined in this manner are still rational; at least this is so in all the examples known to me at present where they can be calculated explicitly. However, the most trivial examples (e. g., that of a quadratic cone in a 3-dimensional projective space) show that such functions need not satisfy a functional equation of the kind which is typical for zeta-functions, so that it is not altogether proper to call them by that name.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] L. Carlitz, "Pairs of quadratic equations in a finite field," *American Journal of Mathematics*, vol. 76 (1954), pp. 137-154.

[2] A. Weil, "Numbers of solutions of equations in finite fields," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 497-508.

ON THE LOCAL BEHAVIOR OF SOLUTIONS OF NON-PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS.

II. The Uniqueness of the Green Singularity.*

By PHILIP HARTMAN and AUREL WINTNER.

---

1. Under the assumption that the coefficients of the partial differential equation

$$u_{xx} + u_{yy} + d(x, y)u_x + e(x, y)u_y + f(x, y)u = 0 \tag{1}$$

are analytic functions of (x, y) for small $x^2 + y^2$, Bôcher [1], pp. 462-464, has shown that if $u = u(x, y)$ is a C^2 (hence, analytic) solution of (1) for small positive $x^2 + y^2$ and satisfies

$$u(x, y) \to \infty \tag{2}$$

as $x^2 + y^2 \to 0$, then

$$u(x, y) \sim \text{Const.} \log (x^2 + y^2)^{-1} \qquad (\text{Const.} > 0). \tag{3}$$

Bôcher's method depends on the use of the Green function belonging to the adjoint of (1). Thus, while his methods are applicable if d, e, f, instead of being analytic, are sufficiently differentiable, they cannot be used if it is only supposed that d, e, f are continuous. It will be shown, however, that his assertion is true under the latter assumption on d, e, f. The proof will depend on a device of Carleman [2], used in [5] for the study of local asymptotic properties of solutions of (1).

It should be pointed out that the proof below is limited to elliptic partial differential equations in two independent variables, while Bôcher's arguments and assertions can be generalized to the case of equations involving n independent variables.

The main theorem to be proved is as follows:

(*) *Let $a(x, y)$, $b(x, y)$, $c(x, y)$, where $ac - b^2 > 0$, be functions of class C^1 on*

$$x^2 + y^2 \leqq R^2, \tag{4}$$

normalized by $a(0, 0) = 1$, $b(0, 0) = 0$, $c(0, 0) = 1$. Let $d(x, y)$, $e(x, y)$, $f(x, y)$

---

* Received December 23, 1953.

be continuous functions on (4). *Let* $u = u(x, y)$ *be a function of class* C^2 *on* $0 < x^2 + y^2 \leqq R^2$ *satisfying*

$$au_{xx} + 2bu_{xy} + cu_{yy} + du_x + eu_y + fu = 0 \tag{5}$$

or, more generally, let $u(x, y)$ *be of class* C^1 *on* $0 < x^2 + y^2 \leqq R^2$ *and satisfy*

$$\int_J (bu_x + cu_y)\,dx - (au_x + bu_y)\,dy = \iint_E (du_x + eu_y + fu)\,dxdy \tag{6}$$

for every simply connected set E *bounded by a piecewise smooth Jordan curve* J *in* $0 < x^2 + y^2 \leqq R^2$. *Finally, as* $x^2 + y^2 \to 0$, *let* (2) *hold. Then* (3) *holds and, in fact, as* $x^2 + y^2 \to 0$,

$$\lim (x + iy)(u_y + iu_x) = c_1 + ic_2 \text{ exists and } c_1 = 0,\ c_2 \neq 0. \tag{7}$$

The assertion (7) means that, if $x = r\cos\phi$, $y = r\sin\phi$,

$$u_x(x, y) = c_2 r^{-1}\cos\phi + o(r^{-1}), \qquad u_y(x, y) = c_2 r^{-1}\sin\phi + o(r^{-1}),$$

as $r \to 0$, uniformly in ϕ. Thus, the quadrature

$$u(x, y) = -\int_r^R (u_x\cos\phi + u_y\sin\phi)\,dr + u(R\cos\phi, R\sin\phi)$$

shows that (3) holds, as $r \to 0$, with Const. $= \frac{1}{2}c_2 (\neq 0)$.

Remark. It will be clear from the proof that (*) remains valid if the expression $du_x + eu_y + fu$ in (5) or (6) is replaced by a continuous function of x, y, u_x, u_y which is majorized by a constant multiple of $|u_x| + |u_y|$ for (x, y) on (4).

2. Theorem (*) complements the following results on solutions $u = u(x, y)$ of (5) or (6): Let the coefficients a, b, c, d, e, f satisfy the conditions of (*) and let $u = u(x, y)$ be a solution of (6) of class C^1 on the disk (4) satisfying

$$u(x, y) = o(r^m), \text{ as } r \to 0, \tag{8_m}$$

for some fixed integer $m = 1, 2, \cdots$. Then

$$|u_x| = o(r^{m-1}), \qquad |u_y| = o(r^{m-1}) \tag{9_m}$$

and, in fact, (7) can be replaced by the statement that

$$\lim (x + iy)^{-m}(u_y + iu_x) \text{ exists} \qquad (r \to 0); \tag{10_m}$$

[5], pp. 449-450. Also, if $u = u(x, y)$ is a solution of class C^1 on $0 < x^2 + y^2 \leqq R^2$ and satisfies (9_m) for some integer $m = 0, -1, -2, \cdots$,

then (10_m) holds. This last assertion was proved in [5], pp. 452-453, for $m = 0$ under the assumption that $a \equiv c \equiv 1$, $b \equiv 0$ and $f \equiv 0$. It is easily seen that the proof (loc. cit.) remains valid for $m = -1, -2, \cdots$; the arguments in [5], pp. 462-464, show that there is no loss of generality in assuming $a \equiv c \equiv 1$, $b \equiv 0$ and the last part of the proof of (*) below shows that the same is true of the assumption $f \equiv 0$.

Curiously, *the assertion that* (10_m) *is a consequence of* (9_m) *is false for every* m $(= 0, \pm 1, \cdots)$ *if the assumption that* a, b, c *in* (6) *are of class* C^1 *is weakened to the hypothesis that* a, b, c *are just continuous (even when* $d \equiv e \equiv f \equiv 0$*)*. In order to see this, consider (for small $x^2 + y^2$) the binary elliptic system

$$v_x = bu_x + cu_y, \qquad v_y = -au_x - bu_y, \tag{11}$$

where

$$a = (1 - y^2F/r^2)(1 - F)^{-\frac{1}{2}}, \qquad b = (xyF/r^2)(1 - F)^{-\frac{1}{2}},$$
$$c = (1 - x^2F/r^2)(1 - F)^{-\frac{1}{2}} \tag{12}$$

and $F = F(r) = 1/\log r$ or $F(r) = 0$ according as $r = (x^2 + y^2)^{\frac{1}{2}} \gtrless 0$. Clearly, (11) is the Cauchy-Riemann system (Beltrami) belonging to the positive definite metric

$$ds^2 = (1 - x^2F/r^2)\,dx^2 - 2(xyF/r^2)\,dxdy + (1 - y^2F/r^2)\,dy^2.$$

The system (11) for the unknowns u, v is equivalent to the integro-differential equation

$$\int_J (bu_x + cu_y)\,dx - (au_x + bu_y)\,dy = 0 \tag{13}$$

for u alone.

It was shown in [3], Appendix, that (11) has no solution $u = u(x, y)$, $v = v(x, y)$ of class C^1 in (4), with arbitrarily small R, for which the Jacobian $\partial(u, v)/\partial(x, y)$ is not 0 at $(x, y) = (0, 0)$; the system (11) was first considered by Lavrentieff [7], p. 420, in a similar connection.

Since $F(r)$ is continuous for $0 \leqq r < 1$ and $F(0) = 0$, the functions a, b, c are continuous on (4) when $R < 1$. Define $\rho = \rho(r)$ by

$$\rho(r) = \exp\left(-\int_r^R r^{-1}(1 - F(r))^{\frac{1}{2}}dr\right) \text{ or } \rho(r) = 0 \text{ according as } r \gtrless 0;$$

so that, by the definition of $F(r)$,

$$\rho(r) \sim \text{const. } r/(-\log r)^{\frac{1}{2}}, \text{ as } r \to 0, \qquad (\text{const.} > 0) \tag{14}$$

and $\rho(r)$ has a continuous derivative given by

(15) $\quad d\rho/dr = r^{-1}\rho(r)\,(1 - F(r))^{\frac{1}{2}}$ or $d\rho/dr = 0$ according as $r \gtrless 0$.

For an integer $k \neq 0$, put

(16_k) $\qquad u(x, y) = \rho^k(r)\cos k\phi, \qquad v(x, y) = -\rho^k(r)\sin k\phi,$

where $x = r\cos\phi$, $y = r\sin\phi$. It follows readily from (15) that if $k = 1, 2, \cdots$, then (16) is a solution of (11) of class C^1 on (4) (correspondingly, $u(x, y)$ is a solution of (13)) and that if $k = -1, -2, \cdots$, then (16) is a solution of (11) of class C^1 on $0 < x^2 + y^2 \leqq R^2$ (correspondingly, $u(x, y)$ is a solution of (13)). It is clear from (14) that the function $u(x, y)$ in (16_k) satisfies (8_m) if $m = k$. It is seen from

$$u_x = kr^{-1}\rho^k[(1 - F)^{\frac{1}{2}}\cos\phi\cos k\phi + \sin\phi\sin k\phi],$$

$$u_y = kr^{-1}\rho^k[(1 - F)^{\frac{1}{2}}\sin\phi\cos k\phi - \cos\phi\sin k\phi]$$

and from (14), that (9_k) also holds; in fact, both u_x and u_y are majorized, as $r \to 0$, by constant multiples of $r^{k-1}/(-\log r)^{\frac{1}{2}k}$. But (10_k) does not hold. This proves the italicized statement for $m = \pm 1, \pm 2, \cdots$.

In order to prove the statement for $m = 0$, let $F(r)$ be replaced by its negative, that is, in (12) and in the definition of $\rho(r)$, let $F(r)$ be $-1/\log r$ or 0 according as $r \gtrless 0$. Then (14) is replaced by

(14 bis) $\qquad \rho(r) \sim \text{const.}\, r(-\log r)^{\frac{1}{2}}$, as $r \to 0$, $\qquad$ (const. > 0)

while the formula in (15) for $d\rho/dr$ remains valid when $r > 0$. The functions (16_k) are of class C^1 and satisfy (11) on $0 < x^2 + y^2 \leqq R^2$ (correspondingly, $u(x, y)$ is a solution of (13)); they are continuous at the origin if $k \geqq 1$ (but are not of class C^1 on (4) unless $k > 1$). The above formulae for u_x, u_y are valid and show that u_x and u_y are majorized by constant multiples of $r^{k-1}(-\log r)^{\frac{1}{2}k}$ and are, therefore, $o(r^{-1})$ as $r \to 0$, if $k = 1$. Thus, the function $u = u(x, y)$ in (16_k) when $k = 1$ satisfies the condition (9_m) corresponding to $m = 0$. On the other hand, (10_0) does not hold. This completes the proof of the last italicized statement.

It remains an open question whether or not (*) is true if the assumption of the C^1-character of a, b, c is weakened to the hypothesis that a, b, c are just continuous. (Note that, in the above examples, $u = -\log\rho(r)$ is a solution of (13) on $0 < x^2 + y^2 \leqq R^2$ satisfying (2), and that (3) and (7) hold.) If (*) is false under the weaker assumption, there remains the question whether or not the "infinity" of u is independent of the solution u satisfying (2).

3. PROOF OF (*): THE CASE $f \equiv 0$. Since $ac - b^2 > 0$ and $a > 0, c > 0$, standard procedures involving the use of conformal coordinates (Lichtenstein) show that there is no loss of generality in supposing that

$$a(x, y) \equiv 1, \qquad b(x, y) \equiv 0, \qquad c(x, y) \equiv 1; \tag{17}$$

cf. [5], pp. 460-464. Following Bôcher [1], pp. 462-464, the proof of (*) will first be given for the case

$$f \equiv 0. \tag{18}$$

The general case will be deduced from the case (18). The identities (17) and (18) reduce (6) to

$$\int_J u_y dx - u_x dy = \iint_E (du_y + eu_x) dx dy. \tag{19}$$

It follows from Corollary 1 of Theorem 1 in [5], pp. 449-450, that the zeros of the gradient of u, that is, the zeros of the complex-valued function

$$w = u_y + iu_x, \tag{20}$$

cannot cluster at any point $(x_0, y_0) \neq (0, 0)$. Note that the condition $u(x_0, y_0) = 0$ of Theorem 1 of [5] need not be satisfied, since the integrand on the right of (19) is a homogeneous linear combination of u_x and u_y (so that (19) is satisfied by the function $u(x, y) - u(x_0, y_0)$). Thus the zeros of w are enumerable, at most. Hence (2) shows that there exist arbitrarily large constants c with the property that $u(x, y) = c$ implies $w(x, y) \neq 0$. It follows that, if such a number c is sufficiently large and if $\epsilon = \epsilon_c > 0$ is sufficiently small, the set of points $z = x + iy$ satisfying $u(x, y) = c$ is on the open ring $\epsilon^2 < x^2 + y^2 < R^2$ and, consequently, consists of one or more disjoint Jordan curves of class C^1. One of these Jordan curves, say $L = L_c$, surrounds the origin, by (2). On the other hand, $u = u(x, y)$ cannot have a local maximum or minimum at any point $(x_0, y_0) \neq (0, 0)$; cf., e. g., [5], pp. 449-450. Hence, the locus $u(x, y) = c$ consists of the Jordan curve $L = L_c$ only.

Clearly, $u(x, y) = c$ on L means that, if $z = x + iy$, the expression

$$wdz = (u_y dx - u_x dy) + i(u_x dx + u_y dy)$$

reduces to $-u_n ds + idu = -u_n ds$, where u_n is the derivative of u in the direction of the inward normal to L and s is the arc-length on J. The assumption (2) implies that $u_n \geqq 0$ on L. Thus, on L,

$$-wdz = u_n ds, \text{ where } u_n \geqq 0. \tag{21}$$

Let S denote the ring-shaped domain bounded by L and the circle $|z| = R$. If (19) is applied to the domain $E = S$, then

$$\int_L u_n ds = -\int_{|z|=R} (u_y dx - u_x dy) + \iint_S W dx dy, \tag{22}$$

where

$$W = du_x + eu_y. \tag{23}$$

Let K denote a constant satisfying

$$|d| \leqq K, \quad |e| \leqq K \quad \text{on} \quad |z| \leqq R. \tag{24}$$

Then, by (20) and (23),

$$|W| \leqq K|w| \quad \text{on} \quad |z| \leqq R. \tag{25}$$

Hence (22) gives

$$\int_L u_n ds \leqq I_1 + KD_1, \tag{26}$$

where

$$I_1 = \int_{|z|=R} |wdz| \text{ and } D_1 = \iint_S |w| dxdy. \tag{27}$$

The identity (19) implies that, in terms of the notations $\zeta = \xi + i\eta$, $w(\zeta) = w(\xi, \eta)$,

$$2\pi i w(\zeta) = \int_{|z|=R} (z-\zeta)^{-1} w dz - \int_L (z-\zeta)^{-1} w dz + \iint_S (z-\zeta)^{-1} W dx dy \tag{28}$$

if ζ is in S; cf. formula (17_k) in [5], p. 455. Hence, (21) and (25) show that

$$2\pi |w(\zeta)| \leqq I_2 + \int_L |z-\zeta|^{-1} u_n ds + KD_2, \tag{29}$$

where

$$I_2 = \int_{|z|=R} |z-\zeta|^{-1} |wdz| \text{ and } D_2 = \iint_S |z-\zeta|^{-1} |w| dxdy. \tag{30}$$

Let (29) be integrated with respect to $d\xi d\eta$ over S. Then the inequality

$$\iint_{|\zeta| \leqq R} |z-\zeta|^{-1} d\xi d\eta \leqq 4\pi R \qquad (|z| \leqq R) \tag{31}$$

shows that $2\pi D_1 \leqq 4\pi R I_1 + 4\pi R \int_L u_n ds + 4\pi K R D_1$. If this is combined with (26), it follows that

$$(2 - 8KR) D_1 \leqq 8RI_1. \tag{32}$$

Let R be so small that

$$4KR < 1. \tag{33}$$

Since I_1 does not depend on the inner boundary L of S, it is seen from (27) and (32) that w is absolutely integrable over $|z| \leqq R$,

$$D < \infty, \text{ where } D = \iint_{|z| \leqq R} |w| \, dxdy. \tag{34}$$

Since the level curve $L = L_c: u(x, y) = c$ shrinks to the origin as $c \to \infty$, an immediate consequence of (22) and (34) is

$$\lim_{c \to \infty} \int_L u_n ds = 2\pi c_2, \tag{35}$$

$$\text{where } 2\pi c_2 = -\int_{|z|=R} (u_y dx - u_x dy) + \iint_{|z| \leqq R} W dxdy.$$

The relation (34) will now be improved to

$$w(\zeta) = O(|\zeta|^{-1}) \text{ as } \zeta \to 0. \tag{36}$$

To this end, let $z = z_0$ be a point of S. If (29) is multiplied by $|\zeta - z_0|^{-1} d\xi d\eta$ and the result integrated over S, it follows from the inequality

$$|\zeta - z_0|^{-1} |z - \zeta|^{-1} \leqq |z - z_0|^{-1} (|z - \zeta|^{-1} + |z_0 - \zeta|^{-1})$$

and from (31) that

$$2\pi D_2 \leqq 8\pi R I_2 + 8\pi R \int_L |z - \zeta|^{-1} u_n ds + 8\pi K R D_2,$$

if z_0 is renamed ζ. In view of (33), this implies that

$$D_2 \leqq \text{const.} \, (I_2 + \int_L |z - \zeta|^{-1} u_n ds);$$

so that, by (29),

$$2\pi |w(\zeta)| \leqq (1 + K \text{ const.}) (I_2 + \int_L |z - \zeta|^{-1} u_n ds).$$

Since $u_n \geqq 0$, there exists a point $z = z^*$ on L with the property that the

last line integral is $|z^*-\zeta|^{-1}$ times the line integral on the left of (26). Hence (26) gives

$$\int_L |z-\zeta|^{-1} u_n ds \leqq |z^*-\zeta|^{-1}(I_1+KD_1),$$

and so

$$2\pi |w(\zeta)| \leqq (1+K\,\text{const.})\{I_2+|z^*-\zeta|^{-1}(I_1+KD_1)\}.$$

If the constant c defining the level curve $L=L_c$: $u(x,y)=c$ tends to ∞, then z^* tends to 0 and D_1 to the finite limit D; cf. (34). Hence the last formula line implies (36).

The estimate (36) and the definition (30) of D_2 show that $D_2 = O(\log|\zeta|^{-1})$ as $\zeta\to 0$. Hence (21), (25) and (28) give

$$2\pi i w(\zeta) = O(1) + \int_L (z-\zeta)^{-1} u_n ds + O(\log|\zeta|^{-1}), \tag{37}$$

as $\zeta\to 0$. Since $u_n \geqq 0$, applications of the mean value theorem to the real and imaginary parts of the last line integral show that there exist two points $z_0=x_0+iy_0$, $z^0=x^0+iy^0$ on L such that the last integral can be written in the form

$$\int_L (z-\zeta)^{-1} u_n ds = \{(x_0-\xi)|z_0-\zeta|^{-2} - i(y^0-\eta)|z^0-\zeta|^{-2}\}\int_L u_n ds.$$

If the constant c defining $L=L_c$: $u(x,y)=c$ tends to ∞, then both points z_0, z^0 of L tend to 0 and so, by (35),

$$\int_L (z-\zeta)^{-1} u_n ds \to -2\pi c_2/\zeta \qquad (c\to\infty). \tag{38}$$

Since $w(\zeta)$ and the constants implicit in the O-terms in (37) are independent of c, it follows from (37) and (38) that, as $\zeta\to 0$,

$$w(\zeta) = ic_2/\zeta + O(\log|\zeta|^{-1}).$$

Consequently, the limit in (7) exists. It is purely imaginary, since the number c_2, defined in (35), is real.

It remains to show that $c_2\neq 0$. If $c_2=0$, then (9) holds with $m=0$, and so (10) holds with $m=0$. In other words, u_x and u_y tend to (finite) limits as $(x,y)\to(0,0)$; in particular, u_x and u_y are bounded near $(x,y)=(0,0)$. But this contradicts the assumption (2). Consequently, $c_2\neq 0$. This proves (*) in the case (18).

4. PROOF OF (*): THE GENERAL CASE. For the general case, the following lemma will be needed:

LEMMA. *If $R > 0$ is sufficiently small, then, under the assumptions of (*) on the functions a, b, c, d, e, f, there exists a function $u = U(x, y)$ of class C^1 on (4) satisfying*

$$U(0, 0) \neq 0 \tag{39}$$

and the relation (6) for every domain E bounded by a piecewise smooth Jordan curve J in (4).

This lemma is contained in [8]. Another method of proof of this lemma will be indicated. It can be supposed, without loss of generality, that (17) holds, so that (6) reduces to

$$\int_J u_y dx - u_x dy = \iint_E (du_x + eu_y + fu)\, dxdy. \tag{40}$$

By standard methods of successive approximations (Korn, Lichtenstein), it can be shown that if $R > 0$ is sufficiently small and if $u(R\cos\phi, R\sin\phi)$ are arbitrarily assigned boundary values of class C^1 on $x^2 + y^2 = R^2$, then the boundary value problem associated with the integro-differential equation (40) has a (unique) solution $u = u(x, y)$ of class C^1 on (4). Furthermore, there exists a constant C with the property that if the 0-th approximation is the harmonic function $u_0(x, y)$ on $x^2 + y^2 < R^2$ which assumes the assigned boundary values on $x^2 + y^2 = R^2$, then $|u(x, y) - u_0(x, y)| \leqq CR$ on (4); the constant C is independent of R and the boundary values $u(R\cos\phi, R\sin\phi)$, if (without loss of generality) the latter do not exceed 1 in absolute value. In particular, if the boundary values are identically 1 (that is, $u(r\cos\phi, R\sin\phi) \equiv 1$), then $u_0(x, y) \equiv 1$ and so $|u(x, y) - 1| \leqq CR$ on (4). If $U(x, y)$ is identified with the corresponding solution $u(x, y)$ and if $R > 0$ is sufficiently small, then it is seen that (39) holds.

The general case of (*) can be deduced from its special case (18) by the variation of constants, employed by Bôcher for the same purpose. Suppose that (6) is of the form (40). For small $x^2 + y^2 \neq 0$, define $v = v(x, y)$ by

$$u(x, y) = U(x, y)v(x, y), \tag{41}$$

where $u = u(x, y)$ is the given solution of (40) satisfying (2) and $u = U(x, y)$ is a solution of (40), as supplied by the Lemma. Then (40) and (41) imply that $v(x, y)$ satisfies

$$\int_J v_y dx - v_x dy = \iint_E (d^* v_x + e^* v_y)\, dxdy, \tag{42}$$

if E is any simply connected domain bounded by a piecewise smooth Jordan curve J in a sufficiently small punctured disk $0 < x^2 + y^2 \leqq \epsilon^2$ and d^*, e^* are the continuous functions $d^* = d + 2U_x/U$, $e^* = e + 2U_y/U$. This is clear if U, v are of class C^2 on $0 < x^2 + y^2 \leqq \epsilon^2$ and otherwise can be deduced from the form of the Green's identity given in the lemma of [6], p. 761.

Since (2) and (39) imply that $|v(x, y)| \to \infty$ as $x^2 + y^2 \to 0$, it follows from (41) and the case (18) of (*) that, as $x^2 + y^2 \to 0$,

$$\lim (x + iy)(v_y + iv_x) = C_1 + iC_2 \text{ exists and } C_1 = 0, C_2 \neq 0 \tag{43}$$

and, therefore, that

$$v(x, y) \sim \text{const. } \log (x^2 + y^2)^{-1} \qquad (\text{const.} \neq 0). \tag{44}$$

Since U is of class C^1 on (4) and since (39) holds, the assertion (7) follows from (41), (43) and (44). This completes the proof of (*).

5. Theorem (*) is, in a certain sense, a uniqueness theorem. The corresponding existence theorem is the following:

(**) *Let a, b, c, d, e, f satisfy the conditions of* (*). *If $R > 0$ is sufficiently small, then there exists a function $u = u(x, y)$ of class C^1 on $0 < x^2 + y^2 \leqq R^2$, satisfying* (2) *and the relation* (6) *for every simply connected region E bounded by a piecewise smooth Jordan curve J in $0 < x^2 + y^2 \leqq R^2$.*

When the coefficients of (5) are analytic, the assertion of (**) is contained in a result proved by Hilbert around 1900 (before the development of his theory of integral equations); cf. [9], p. 570 and [4], pp. 100-101.

Theorem (**) is contained in a result of Lichtenstein [8] dealing with the existence of a Green's function for the integro-differential equation (6). His proof depends on the theory of integral equations. The assertion (**) can also be proved by the method of successive approximations, after an application of a variation of constants, as follows:

Let (6) be of the form (40) and define a new dependent variable $v = v(x, y)$ by

$$u = v \log (x^2 + y^2).$$

Then the identity (40) is equivalent to the identity

$$\int_J v_y dx - v_x dy = \iint_E (d^* v_x + e^* v_y + f^* v)\, dx dy, \tag{45}$$

where

$$d^* = d + 4x/\{(x^2+y^2)\log(x^2+y^2)\},\quad e^* = e + 4y/\{(x^2+y^2)\log(x^2+y^2)\},$$
$$f^* = f + 2(xd+ye)/\{(x^2+y^2)\log(x^2+y^2)\}.$$

The fact that d^*, e^*, f^* are majorized by Const./$\{(x^2+y^2)^{\frac{1}{2}}\log(x^2+y^2)^{-1}\}$ makes possible the application of the standard method of successive approximations to (45), proving that if $R>0$ is sufficiently small and if $v(R\cos\phi, R\sin\phi)$ are arbitrarily assigned boundary values of class C^1 on $x^2+y^2=R^2$, then the boundary value problem associated with (45) has a solution $v=v(x,y)$ which is continuous on (4) and of class C^1 on $0<x^2+y^2\leqq R^2$ (in fact, such that $v_x/\log(x^2+y^2)$ and $v_y/\log(x^2+y^2)$ are bounded as $x^2+y^2\to 0$). Finally, as in the proof the Lemma above, $v(0,0)\neq 0$ if R is sufficiently small and the boundary conditions are suitably chosen, for example, if $v(R\cos\phi, R\sin\phi)\equiv 1$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, "Singular points of functions which satisfy partial differential equations of the elliptic type," *Bulletin of the American Mathematical Society*, vol. 9 (1903), pp. 455-465.

[2] T. Carleman, "Sur les systèmes linéaires aux dérivées partielles du premier ordre à deux variables," *Comptes Rendus*, vol. 197 (1933), pp. 471-474.

[3] S. S. Chern, P. Hartman and A. Wintner, "On isothermic coordinates," to appear in the *Commentarii Mathematici Helvetici*.

[4] J. Hadamard, "Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques," Paris, 1932.

[5] P. Hartman and A. Wintner, "On the local behavior of solutions of non-parabolic partial differential equations," *American Journal of Mathematics*, vol. 75 (1953), pp. 449-476.

[6] ——— and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[7] M. Lavrentieff, "Sur une classe de représentations continues," *Recueil Mathématique*, vol. 42 (1935), pp. 407-424.

[8] L. Lichtenstein, "Neue Untersuchungen in der Theorie der linearen partiellen Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Bulletin de l'Académie Polonaise des Sciences et des Lettres* (1931), pp. 571-598.

[9] A. Sommerfeld, Article II A7c (1900) in the *Encyklopädie der mathematischen Wissenschaften*.

SOME APPLICATIONS OF THE LOCAL UNIFORMIZATION THEOREM.*

By SERGE LANG.

1. Introduction. Let K be a function field over a constant field k. We shall always assume that K/k is regular: k is algebraically closed in K and K/k is separably generated. If $K = k(x_1, \cdots, x_n) = k(x)$, we may view (x) as the generic point (over k) of a variety V, which is then called a model of K. If $\mathfrak{p}$ is a place of K/k, i. e. a place which is an isomorphism, or identity, on k, and if we put $\bar{x}_i = x_i(\mathfrak{p}) \neq \infty$, then $(\bar{x})$ is a point of V and we say that $(\bar{x})$ is at the center of $\mathfrak{p}$.

Zariski's Theorem of Local Uniformization [10] states that given a place $\mathfrak{p}$ of K/k (characteristic zero) it is always possible to find a model of the field for which the point at the center of $\mathfrak{p}$ is non singular.

In the classical case where k is the complex numbers, this implies that the place has a certain neighborhood.

It is our purpose to show that it is not the algebraic closure of the complex numbers which is essential for the existence of this neighborhood, but rather its topological completeness.

More precisely: Suppose that the constant field k of K is complete under a valuation, and of characteristic zero. We say that a place of K/k is rational over k if it is k-valued. Let $\mathfrak{M}$ consist of all rational places of K/k, and assume that $\mathfrak{M}$ is not empty. $\mathfrak{M}$ may be called the rational Riemann manifold of K/k, but as we shall deal only with rational places, we shall call it the Riemann manifold. We topologize $\mathfrak{M}$ (as it is done classically) by giving it the weakest topology such that the map $x \to x(\mathfrak{p}) \in k_\infty$ is continuous for each x in K.

One can then prove that $\mathfrak{M}$ contains infinitely many places, and in fact enough places to distinguish between elements of K. If k is locally compact, then $\mathfrak{M}$ is a compact space, and in the case where K has dimension 1 over k (the case of curves) then $\mathfrak{M}$ is locally homeomorphic to k itself. Our proofs are closely related to Chevalley's [2].

Kawada [4] has also considered the abstract manifold in the case of curves, but restricted himself to the case where k is algebraically closed. In

* Received March 16, 1953.

that case, the complex numbers are the only locally compact field. If however we do not assume k algebraically closed, then we can include the real numbers, and p-adic fields. In fact, in the case of curves, we can also treat the power series over finite fields because the uniformization theorem is true irrespective of the characteristic.

If k is the real numbers, then $\mathfrak{M}$ consists of a finite number of components, each homeomorphic to a circle. It is a classical theorem of Harnack [3] that the number of these components is at most equal to $g+1$, where g is the genus of K. We shall give a simple new proof of this theorem, based on the Riemann-Roch Theorem.

2. The Riemann manifold. We let k be a field complete under a rank one valuation, and do not make any restrictions on the characteristic. We view the value group as a subgroup of the positive reals. If $a \in k$, we let $|a|$ denote its value.

We suppose that K/k is a function field with the properties stated in the introduction. We assume that there exists at least one rational place of K/k, and let $\mathfrak{M}$ be the set of all such places. For each $x \in K$, we let Γ_x be the "k-sphere" k_∞ obtained from k by the formal adjunction of ∞, and topologized in the obvious way: Neighborhoods of ∞ are the outsides of circles around 0. We let $\Gamma = \prod_x \Gamma_x$ be the Cartesian product taken over all $x \in K$, with the product topology.

Let $\mathfrak{p} \in \mathfrak{M}$. Then $\mathfrak{p}$ maps K into k_∞. Let Φ be the map of $\mathfrak{M}$ into Γ defined by $\Phi : \mathfrak{p} \to (\cdots, x(\mathfrak{p}), \cdots)$ and denote $\Phi(\mathfrak{M})$ by $\mathfrak{M}'$. Then Φ is obviously 1-1.

We topologize $\mathfrak{M}'$ as a subspace of Γ, and give $\mathfrak{M}$ the weakest topology so that Φ is continuous. In other words, the open sets of $\mathfrak{M}$ consist of the sets $\Phi^{-1}(U)$ where U is open in $\mathfrak{M}'$

The following theorem and its proof are identical with Chevalley's [2], except for the fact that only the completeness of k is used, and not its algebraic closure.

THEOREM 1. *$\mathfrak{M}'$ is closed in Γ.*

Proof. Let $u = (\cdots, u_x, \cdots)$ be in the closure of $\mathfrak{M}'$. Let $\mathfrak{o} = \{x \in K \mid u_x \neq \infty\}$. We shall prove that $\mathfrak{o}$ is a valuation ring (or v-ring) whose induced place $\mathfrak{p}$ is such that $x(\mathfrak{p}) = u_x$.

We first note that $k \subset \mathfrak{o}$, because $a(\mathfrak{p}) = a$ for all $a \in k$, and all $\mathfrak{p} \in \mathfrak{M}$.

Let x and y lie in $\mathfrak{o}$. Then $u_x, u_y \neq \infty$. Since u is a limit point of

elements in $\mathfrak{M}'$, there exists a prime $\mathfrak{p} \in \mathfrak{M}$ such that $x(\mathfrak{p})$, $y(\mathfrak{p})$, and $(x+y)(\mathfrak{p})$ are arbitrarily close to u_x, u_y, and u_{x+y} respectively. For such $\mathfrak{p}$, we see that $x(\mathfrak{p})$ and $y(\mathfrak{p}) \neq \infty$, whence $(x+y)(\mathfrak{p}) = x(\mathfrak{p}) + y(\mathfrak{p}) \neq \infty$. Hence $x+y \in \mathfrak{o}$. Similarly, $x-y$ and xy lie in $\mathfrak{o}$, which is therefore a ring.

Furthermore, the map $x \to u_x$ is a homomorphism of $\mathfrak{o}$ into k and is identity on k. This follows from a continuity argument as above.

Finally, $\mathfrak{o}$ is a v-ring. Suppose $x \notin \mathfrak{o}$, Then $u_x = \infty$. Let $y = x^{-1}$. There exists a prime $\mathfrak{p} \in \mathfrak{M}$ such that $x(\mathfrak{p})$ is close to u_x and $y(\mathfrak{p})$ is close to u_y. But $x(\mathfrak{p})$ close to u_x means $|x(\mathfrak{p})|$ is very large. Hence $|y(\mathfrak{p})|$ is very small. Hence $u_y = 0$, so $y \in \mathfrak{o}$. Thus $\mathfrak{o}$ is a v-ring.

Since $x \to u_x$ is a homomorphism of $\mathfrak{o}$ into k, and since $\mathfrak{o}$ is a v-ring, it follows that $x \to u_x$ is a place of K, which is rational because $u_x \in k_\infty$. Our theorem is thereby proved.

In particular, if k is locally compact, then k_∞ is the ordinary one-point compactification, and k_∞ is therefore compact. By Tychonoff's theorem, Γ is compact. Since $\mathfrak{M}'$ is closed in Γ, it is also compact. We obtain therefore the following.

COROLLARY. *If k is locally compact, then $\mathfrak{M}$ is compact.*

We shall now investigate more closely the special one-dimensional case. For the rest of this section, we suppose that K/k has transcendence degree 1, and is regular. For the convenience of the reader, we reproduce a proof of the local uniformization theorem.

Let $\mathfrak{p}$ be a prime of K, whose residue class field is separable over k. Let $x \in K$ be a separating variable, such that $\operatorname{ord}_{\mathfrak{p}} x = 1$. (It is well known that such an x can always be found. See for instance [1], Ch. 17, 4.) If p is the rational prime induced by $\mathfrak{p}$ on $k(x)$, then p is unramified in K. Hence there exists a generator y of K over $k(x)$ which is integral at $\mathfrak{p}$ and whose local $\mathfrak{p}$-different is not $\equiv 0 \pmod{\mathfrak{p}}$. This means: If $f(x,y) = 0$ is the equation for x and y over k, and if we put $\bar{x} = x(\mathfrak{p})$, $\bar{y} = y(\mathfrak{p})$ then $f_y(\bar{x}, \bar{y}) \neq 0$. Hence $(\bar{x}, \bar{y})$ is non-singular at the center of $\mathfrak{p}$ as was to be shown.

We recall one more useful theorem from the theory of curves.

Let $K = k(x,y)$ be the function field of a curve, and let $f(x,y) = 0$ be its equation over k. Let $(\bar{x}, \bar{y})$ be a non-singular point on the curve, so that $f(\bar{x}, \bar{y}) = 0$ but $f_y(\bar{x}, \bar{y}) \neq 0$. Then $(\bar{x}, \bar{y})$ is at the center of exactly one place of K/k. In other words, there exists one and only place $\mathfrak{p}$ of K/k such that $\bar{x} = x(\mathfrak{p})$, $\bar{y} = y(\mathfrak{p})$. Furthermore, the residue class field $\bar{K}$ of K under $\mathfrak{p}$ is naturally isomorphic to $k(\bar{x}, \bar{y})$.

For a proof, see Artin [1], Ch. 16, 2.

In particular, if $(\bar{x}, \bar{y})$ is a non-singular rational point (that is if its coordinates lie in k), then the place above it is also rational.

Let $\mathfrak{p}$ be a given prime of $\mathfrak{M}$. We shall find a neighborhood for $\mathfrak{p}$. Let x be a local uniformizing parameter at $\mathfrak{p}$, separating for K/k. In other words, assume $\operatorname{ord}_{\mathfrak{p}} x = 1$, and $K/k(x)$ is separable. According to the local uniformizing theorem, we can find a model $K = k(x, y)$ such that the point $(\bar{x}, \bar{y})$ at the center of $\mathfrak{p}$ is non-singular: $f_y(\bar{x}, \bar{y}) \neq 0$. By multiplying x and y by a constant $a \in k$ having sufficiently small value, we may assume that $\bar{x}$ and $\bar{y}$ are integral in k. This will not affect the non-singularity of the point. We may also assume that f has integral coefficients by clearing denominators if necessary. These preparations are necessary in order to be able to apply the Newton approximation method later. (Cf. [5].) They apply of course only if the valuation of k is non-archimedean.

Split the polynomial $f(\bar{x}, Y)$ in the algebraic closure $\bar{k}$ of k. Then $\bar{y}$ is a root of multiplicity 1, so we have $f(\bar{x}, Y) = (Y - \bar{y})(Y - y_1)^{e_1} \cdots (Y - y_r)^{e_r}$. We recall that the roots of an algebraic equation are continuous functions of the coefficients ([1], Ch. 2, 6). Given a neighborhood $\bar{M}$ of $\bar{y}$ in $\bar{k}$ not containing $y_1, \cdots, y_r$ there exists a neighborhood N_1 of x in k such that for all values $\xi \in N_1$, the equation $f(\xi, Y) = 0$ has exactly one root in M.

Let $M = \bar{M} \cap k$. There exists a neighborhood $N \subset N_1$ of x such that for any value $\xi \in N$, we can refine $\bar{y}$ to a root η of $f(\xi, Y)$ with $\eta \in M$. This can be done with the Newton approximation method, as given in [5]. Since the partial derivative is continuous, we also have $f_y(\xi, \eta) \neq 0$. Hence there exists a unique rational place $\mathfrak{q}$ of K such that $x(\mathfrak{q}) = \xi$ and $y(\mathfrak{q}) = \eta$. This proves that the map $\mathfrak{q} \to x(\mathfrak{q})$ is 1-1 for all $\mathfrak{q}$ such that $(x(\mathfrak{q}), y(\mathfrak{q}))$ lies in the neighborhood (N, M) of $(\bar{x}, \bar{y})$. By what has been said above, this latter condition is equivalent to the single condition that $x(\mathfrak{q}) \in N$.

The map is continuous. We may select N and M to be closed "discs" around $\bar{x}$ and $\bar{y}$ respectively, in which case the set of primes $\mathfrak{q}$ mapping x and y into (N, M) is closed.

If k is locally compact, this set of primes is compact. Hence the map is a homeomorphism. Summarizing, we have

THEOREM 2. *If k is locally compact, and if K/k is a regular extension of dimension one, then $\mathfrak{M}$ is locally homeomorphic with k. Given a prime $\mathfrak{p} \in \mathfrak{M}$, and a separating local uniformizing parameter $x \in K$ at $\mathfrak{p}$, the map $\mathfrak{q} \to x(\mathfrak{q})$ gives a homeomorphism of a certain neighborhood of $\mathfrak{p}$ with a disc of center 0 in k.*

If $k = R$ is the real numbers, then $\mathfrak{M}$ is the union of a finite number of components, which have classically been called the real branches of the curve. It is obvious (and in fact can easily be proved) that each component of $\mathfrak{M}$ is homeomorphic to a circle. (One uses the facts that it is compact, connected, and locally homeomorphic to real segments.) Harnack [3] has proved that the number of these branches cannot exceed $g + 1$, where g is the genus of the curve. We give below a proof depending on the Riemann-Roch theorem.

We begin by a remark. Let B be one of the branches, and let $x \in K$ be a function having no pole on B. We assert that x cannot have only a single zero on B. In other words, x has either no zero on B, or has at least two zeros, counting multiplicities. Indeed, if x has only a single zero $\mathfrak{p}$, then $x(\mathfrak{q}) > 0$ and $x(\mathfrak{q}') < 0$ for $\mathfrak{q}$ and $\mathfrak{q}'$ in a small neighborhood of $\mathfrak{p}$. Naively speaking, x changes sign as it runs through $\mathfrak{p}$. But $B - \{\mathfrak{p}\}$ is connected. Since x gives a continuous function of $B - \{\mathfrak{p}\}$, it must therefore adopt the value 0 somewhere in $B - \{\mathfrak{p}\}$, contradiction.

Now suppose there are $g + 2$ branches. Select one place on each branch, and let $\mathfrak{p}_1, \cdots, \mathfrak{p}_{g+2}$ be these places. Let $\mathfrak{q}$ be a place of K of degree 2 (a complex place). Then $\mathfrak{q}$ does not lie on any one of the real branches. The divisor $\mathfrak{p}_1 \cdots \mathfrak{p}_{g+2}/\mathfrak{q}^{g+1}$ has degree $g + 2 - 2(g + 1) = -g$. By the Riemann-Roch theorem, there exists a function $x \in K$ which is multiple of this divisor. Thus x has g more zeros in addition to $\mathfrak{p}_1, \cdots, \mathfrak{p}_{g+2}$, and has no pole on any of the real branches. This implies that on at least two branches, x will have only a single zero, a contradiction.

Harnack proved his theorem by taking a model of the curve, and assuming that the singularities were not too horrible. The classical proof is obtained by viewing the real branches as lying on the complex Riemann surface, by cutting the latter along these real branches, and using a symmetry argument: the real branches consist precisely of those places on the surface which are invariant under the complex conjugate automorphism.

Our proof is of course much more algebraic. In fact, it is very likely that one can define branches for a curve over an arbitrary real closed field, and that one can prove in that case also that there are at most $g + 1$ branches. As we have not been able to carry out all the details of this idea, we shall leave it as it stands for the moment.

If one takes a model for the curve, $K = k(x, y)$, then it is extremely difficult, if not impossible, to give a rigorous definition of a branch in the (x, y) plane without leaving the plane. For instance, we want to say that a lemniscate is one branch, but that two tangent circles are two branches.

The correct way of defining a branch in the (x, y) plane is to let it consist of the projection on this plane of one of the components B of $\mathfrak{M}$. This projection will be the set of all points $(x(\mathfrak{p}), y(\mathfrak{p}))$ with $\mathfrak{p} \in B$. We note that this projection on the plane may not be connected (as in the case of a hyperbola, where the connection is at infinity), and may also intersect the projection of another branch. On the real manifold $\mathfrak{M}$, however, two branches can never intersect.

3. Function fields over complete fields. In this section, we return to the general case, but assume that k has characteristic zero. It is again complete under a rank one valuation. We assume that k is contained in a universal domain of infinite degree of transcendence over k, and in general follow the conventions of Weil [8]. However, we do not distinguish between fields and abstract fields, as the context will always make our meaning clear.

We say that an extension E/F of a field F is regular if every finitely generated intermediate field E' with $F \subset E' \subset E$ is regular over F.

A subfield k_1 of k will be called admissible if it is dense in k, if k/k_1 is regular, and if k/k_1 has infinite degree of transcendence.

Our main result is concerned with function fields over admissible fields.

THEOREM 3. *Let k_1 be admissible. Let K/k_1 be a finitely generated extension, regular over k_1. There exists a rational place of K/k_1 if and only if there exists an isomorphism of K into k which is identity on k_1. If this is the case, then there exists infinitely many rational places. In fact, given a finite set of quantities $(z) = (z_1, \cdots, z_n)$ with $z_i \in K$, and a finite set of polynomials $g_j(Z) \in k_1[Z]$ such that $g_j(z) \neq 0$, there exists a rational place $\mathfrak{p}$ of K/k_1 such that all $\bar{z}_i = z_i(\mathfrak{p})$ are finite and such that $g_j(\bar{z}) \neq 0$.*

Proof. Suppose that there exists a rational place $\mathfrak{p}$ of K/k_1. By the local uniformization theorem, we can find a model of K whose center at $\mathfrak{p}$ is non-singular. By a sufficiently general projection, we may assume that this model V is a non-singular hypersurface: $K = k_1(x_1, \cdots, x_r, y)$ where (x, y) satisfy the irreducible equation $f(x, y) = 0$ over $k_1, x_1, \cdots, x_r$ are algebraically independent over k_1, and since V is non-singular, $f_y(\bar{x}, \bar{y}) \neq 0$.

By multiplying if necessary the elements x_i, y by an element $a \in k_1$ having sufficiently small value, we may assume that $\bar{x}_i$, $\bar{y}$ are integral in k_1. Namely we have $(ax)(\mathfrak{p}) = a \cdot x(\mathfrak{p})$ because k_1 is fixed under $\mathfrak{p}$. Such a transformation does not affect the non-singularity of the point. In addition we may assume that f has integral coefficients, by clearing its denominators if necessary.

We can now apply the Newton approximation method as in [5]. Select $\xi_1, \cdots, \xi_r$ very close to $x_1, \cdots, x_r$ in k, but algebraically independent over k_1. Our assumptions on k allow us to do this. Then y can be refined to a root η of $f(\xi, Y)$ in k, and it is clear that $k_1(\xi, \eta)$ is isomorphic to $k_1(x, y)$. This proves the first part of our theorem. We note that we have a tremendous amount of freedom in the isomorphism of K into k.

Assume now that $k_1 \subset K \subset k$. We must show that there exists rational places of K/k_1. This will be done by an induction on the transcendence degree of K over k_1. If this transcendence degree is 1, let $K = k_1(x, y)$, and let $f(x, y) = 0$ be the irreducible equation for x and y over k_1. Then $f_y(x, y) \neq 0$ because (x, y) is generic. As before we may assume that x and y are integral in k and that f has integral coefficients. We then use the Newton method in the reverse direction. We can select ξ in k_1 arbitrarily close to x so that y can be refined to a root of $f(\xi, Y)$. This root will lie in k, and will be algebraic over k_1, hence in k_1. This procedure yields infinitely many points on the curve $f(x, y) = 0$, and almost all of them will be non-singular. We had pointed out that a non-singular point of a curve lies at the center of exactly one place and that the residue class field of this place is then generated by the coordinates of the point over the constant field. Since the coordinates obtained in our proof are rational, we have proved what we wanted.

Our construction shows in addition that any finite set of quantities (z) will not be mapped into 0 or ∞ by almost all the places. In particular, we can find places such that none of the quantities $g_j(z)$ are mapped into 0 or ∞. Furthermore, if we view (z) as a generic point, infinitely many of the places just constructed will induce specializations of (z) into k_1 which are arbitrarily close to (z) itself. For a proof and an application of this, see [5].

It is now easy to complete the induction. Let $k_1 \subset E \subset K$, where K/E has dimension 1. Let E_1 be the algebraic closure of E in k. Then E_1 is admissible. Let $L = KE_1$. Then L may be viewed as a function field in one variable over the constant field E_1. There exists a place $\mathfrak{p}$ of L rational over E_1, and $\mathfrak{p}$ induces a place on K/E which is E_1-valued. Since K/E has dimension 1, the image $\bar{K}$ of K under $\mathfrak{p}$ is finitely generated over E_1, and consequently over k_1. Since $\bar{K}$ has smaller dimension over k_1 than K we conclude by induction that there exists a place $\mathfrak{q}$ of $\bar{K}$ which is rational over k_1, and satisfying all the extra algebraic requirements that we desire. The composite of the places $\mathfrak{p}$ and $\mathfrak{q}$ is a rational place of K over k_1 satisfying all our conditions. This concludes our proof.

Theorem 2 can be used to study function fields over admissible fields k_1

which are interesting in themselves, such as the maximal unramified extension of a p-adic field, or the convergent power series over a valuated field. It can also be extended to function fields over k itself. The fields k_1 then play the role of fields of definition. This is done as follows.

Let K/k be a function field, and assume that there exists one rational place $\mathfrak{p}$ of K/k. Let V be a hypersurface which is a model of K and is nonsingular at the center of $\mathfrak{p}$. Let $K = k(x)$ where (x) is a generic point of V. Let k_1 be an admissible subfield of k which contains a field of definition for V (i. e. all the coefficients of the irreducible polynomial $f(X)$ vanishing on (x)), and all the values $x_i(\mathfrak{p})$. Such a field may be obtained by adjoining to an admissible field the finite set of quantities just mentioned, and taking the relative algebraic closure in k.

Let $K_1 = k_1(x)$. By a procedure similar to the one carried out in the first part of the proof of Theorem 3, there exists an isomorphism of K_1 into k which is identity on k_1. Hence we can find infinitely many places $\mathfrak{q}$ of K_1 which are rational over k_1. All that remains to be shown is that such places extend in a natural way to rational places of K over k.

The field K_1 is linearly disjoint from k over k_1 by Weil [8], Ch. I, 6, Th. 3. Our statement now follows from the following general lemma.

LEMMA. *Let K, L be two extensions of a field F and suppose that they are linearly disjoint over F. Let $\mathfrak{q}$ be a rational place of K/F. Then $\mathfrak{q}$ extends uniquely to a rational place of KL/L.*

Proof. Let $K_\mathfrak{q}$ be the ring of elements of K which are finite under $\mathfrak{q}$. The specialization on $K_\mathfrak{q}$ induced by $\mathfrak{q}$ extends uniquely to a specialization of the ring $L[K_\mathfrak{q}]$ generated over L by the elements of $K_\mathfrak{q}$. This is well known for a finite set of quantities, and the generalization to an infinite set is immediate.

Extend this specialization to a place $\mathfrak{q}$ of KL over L. We contend that it is uniquely determined, and that it is L-valued.

An element of KL can be written

$$\gamma = (a_1\alpha_1 + \cdots + a_r\alpha_r)/(b_1\beta_1 + \cdots + b_s\beta_s)$$

with $a_i, b_j \in K$, and $\alpha_i, \beta_j \in L$, and with the β_j linearly independent over F.

Say b_1 has maximal value under $\mathfrak{q}$, or equivalently under $\mathfrak{Q}$. Divide numerator and denominator by b_1. We get

$$\gamma = ((a_1/b_1)\alpha_1 + \cdots + (a_r/b_1)\alpha_r)/(\beta_1 + \cdots + (b_s/b_1)\beta_s)$$

where each a_i/b_1 and b_j/b_1 is finite under $\mathfrak{Q}$, and lies in F. Since $\mathfrak{Q}$ is identity

on L, and since the β_j are linearly independent over F, the denominator is now finite under $\mathfrak{Q}$, and is not $\equiv 0 \pmod{\mathfrak{Q}}$. The numerator is finite, and clearly both are L-valued. This concludes our proof.

We may summarize the preceding results:

THEOREM 4. *Let K/k be a function field over a field k complete under a valuation, and of characteristic zero. Assume that there exists one rational place of K/k. Then there exists infinitely many such places. In fact, given a finite set of algebraic conditions as in Theorem* 3, *we can find a rational place satisfying these conditions.*

In order to take care of the algebraic conditions on a finite set of elements (z) in K, we must be sure that all the elements of this set will lie in the field K_1. As they can be expressed as rational functions of (x) with a finite number of coefficients from k, all we need to do is to take a field k_1 containing these coefficients.

We shall now make a few additional remarks on function fields over complete fields, which admit rational places.

If we work over the complete field k itself, there is in general no natural extension of k in which we can imbed the function field K having a rational place, as we could do if we worked over an admissible subfield. However, there is an interesting special case where such an extension exists.

Namely, let k be complete under a discrete valuation, whose residue class field is algebraically closed. Then k can be represented as a power series field or a field of Witt vectors [9] with components in a field $\mathfrak{k}$ isomorphic to the residue class field. Denote such a complete field by $F(\mathfrak{k})$. Let $\mathfrak{K}$ be an algebraically closed field containing $\mathfrak{k}$ and assume $\mathfrak{K}/\mathfrak{k}$ has infinite degree of transcendence. Then $F(\mathfrak{k})$ is contained naturally in $F(\mathfrak{K})$. It turns out that $F(\mathfrak{K})$ plays towards $F(\mathfrak{k})$ the same role that k played towards k_1 in Theorem 4. In fact, one can replace k and k_1 by $F(\mathfrak{K})$ and $F(\mathfrak{k})$ everywhere in Theorem 4 without losing the validity of the theorem. The proof is entirely similar, and we shall not reproduce the details of the argument (cf. [5]).

The results which have just been obtained may be interpreted as follows: A complete field may be viewed as a sort of universal domain for those function fields over it which admit a rational place.

The admissible subfields behave as the natural fields of definition for varieties defined over k. However, the restriction to admissible subfields was partly done for convenience. For instance, if a function field K over the rational numbers Q has a rational place, it is clear from our theorems that

there exists an isomorphism of K/Q into each p-adic field, and into the real numbers.

4. Additional remarks. It is perhaps worth while to note the analogy of Theorem 3 with the isomorphism theorem for real function fields K over real closed fields P, as proved in [6]. In the latter case, the internal criterion of reality suffices to yield the existence of a real, zero-dimensional place, and to prove that K can be mapped isomorphically into every real closed field of sufficiently high transcendence degree over P. In the case of complete fields, no such internal criterion seems to be available: One has to assume a priori the existence of at least one rational place in order to get enough of them. It would of course be of great interest to find such an internal criterion, if it exists.

Addendum.* We take this opportunity to give another proof of the theorem for real function fields, following as closely as possible the proof of Theorem 3.

THEOREM 5. *Let K be a real function field over a real closed constant field* P. *Let $z_j \in K$, $z_j \neq 0$ $(j = 1, \cdots, m)$ be a finite set of functions of K. Then there exists a rational place $\mathfrak{p}$ of K/P such that each $z_j(\mathfrak{p})$ is finite and $\neq 0$.*

Proof. By essentially the same inductive step used in Theorem 3, and in the corresponding theorem of [6], we can assume that K/P has transcendence degree 1. It suffices then to prove that for a suitable model of K, we can find infinitely many points on the curve with coordinates in P. We proceed to do this. We suppose that K is ordered, and distinguish two cases.

Suppose first the ordering of K/P is non archimedean. Then the canonical place $\mathfrak{p}$ of K/P is rational over P. Let $\mathrm{P}(x, y)$ be a non-singular model of K at the center of $\mathfrak{p}$. Then $f(\bar{x}, \bar{y}) = 0$ but $f_y(\bar{x}, \bar{y}) \neq 0$ and hence for small δ in P $f(\bar{x}, \bar{y} + \delta) = \delta f_y(\bar{x}, \bar{y}) + \alpha$ where α is much smaller than δ. This means that $f(\bar{x}, Y)$ changes sign at $\bar{y} + \delta$ and $\bar{y} - \delta$. For all $\xi \in \mathrm{P}$ sufficiently close to $\bar{x}$, $f(\xi, Y)$ will likewise change sign at $\bar{y} + \delta$ and $\bar{y} - \delta$. It will therefore have a zero in P, and we get in this way infinitely many zeros of $f(X, Y)$, as desired.

If the ordering of K/P is archimedean, we take any model $K = \mathrm{P}(x, y)$.

* Received April 9, 1953.

Then $f_y(x, y) \neq 0$ because (x, y) is generic. The same procedure as above can now be applied to get the rational points in this case. This proves our theorem.

The isomorphism theorem can also be obtained in a similar way:

THEOREM 6. *Let $K/\mathbf{P}$ be a real function field of transcendence degree r. Let Γ be a real closed field containing $\mathbf{P}$, and having transcendence degree $\geq r$ over $\mathbf{P}$. Then there is an isomorphism of K into Γ which is identity on $\mathbf{P}$.*

Proof. Let $K = \mathbf{P}(x_1, \cdots, x_r, y)$ where (x, y) satisfies over $\mathbf{P}$ an irreducible equation $f(x, y) = 0$, $f_y(x, y) \neq 0$. Using Theorem 5 we can find a rational place $\mathfrak{p}$ of $K/\mathbf{P}$ such that the point $(\bar{x}, \bar{y})$ at the center of $\mathfrak{p}$ is non-singular: $f(\bar{x}, \bar{y}) = 0$ but $f_y(\bar{x}, \bar{y}) \neq 0$. This means that $f(\bar{x}, Y)$ changes sign in $\mathbf{P}$ around $\bar{y}$. For a point (ξ) with coordinates in Γ sufficiently close to those of $(\bar{x})$, the polynomial $f(\xi, Y)$ will change sign in Γ. We may clearly select (ξ) algebraically independent over $\mathbf{P}$. If $f(\xi, \eta) = 0$ for $\eta \in \Gamma$, we see that $\mathbf{P}(\xi, \eta)$ is isomorphic to $\mathbf{P}(x, y)$.

The proofs of Theorems 3 and 5 are essentially alike, differing only in that the existence of rational points on a curve in the neighborhood of a given non-singular point obtained in one case by the Newton method, is obtained in the case of real closed fields by using the intermediate value theorem. This procedure avoids Sturm's Theorem, and uses instead a more naive property of real closed fields. It also has the advantage of unifying the local techniques used in both cases, so that the ordinary real numbers appear as a common field for both cases. It must be remarked finally that although the full force of the local uniformization theorem was used in Theorem 3, only the special case of curves was needed to handle the real fields. This means that the proofs of Theorems 5 and 6 remain elementary.

An interesting possibility for generalization lies in the following direction: How far do the results concerning specializations remain valid if one replaces the local fields by the quotient fields of complete local domains? The trouble arising here is partly due to the fact that the topology on the local domains has no natural extension to its quotient field, and consequently makes the handling of specializations over such fields more difficult. In particular, we know of no analogue to the Newton approximation method.

One might for instance make the following conjecture: Let $\mathfrak{o}_1$ be the ring of convergent power series in several variables over the complex numbers, and let $\mathfrak{o}$ be the ring of formal power series. Let k_1 and k be the respective

quotient fields. Let $(x) = (x_1, \cdots, x_n)$ be a generic point over k_1, with components in $\mathfrak{o}$. Does there exist a specialization of (x) into $\mathfrak{o}_1$?

It is actually possible to prove the following result. Let $\mathfrak{o}_1 = \mathfrak{k}\{t\}$ be the ring of power series in several variables, over an algebraically closed field $\mathfrak{k}$ whose cardinality is greater than denumerable. Let $\mathfrak{K}$ be an algebraically closed field containing $\mathfrak{k}$, and let $\mathfrak{o} = \mathfrak{K}\{t\}$ be the power series over $\mathfrak{K}$. Then $\mathfrak{o}_1 \subset \mathfrak{o}$ in a natural way. Let F_1 and F be the respective quotient fields.

Let $x_i(t) \in \mathfrak{K}\{t\}$ be a finite set of power series, and put $(x) = (x_1, \cdots, x_n)$. View (x) as a generic point over F_1. Then (x) has a specialization $(\bar{x})$ over F_1, with $\bar{x}_i \in \mathfrak{o}_1$.

Indeed, let $f_j(X)$ be a basis for the ideal in $F_1[X]$ vanishing on (x). By clearing denominators, we may assume that all coefficients of f_j lie in $\mathfrak{o}_1$.

Write each variable X_i as a power series $X_i(t)$ with indeterminate coefficients: $X_i = \sum_{\mu} \xi_{i\mu}\mu$ where μ ranges over all monomials $t_1^{\mu_1} \cdots t_s^{\mu_s}$ in t's. Then we can write $f_j(X) = \sum_{\mu} f_{j\mu}(\xi)\mu$ where $f_{j\mu}(\xi) \in \mathfrak{k}[\xi]$. That all f_j vanish on (x) implies that the denumerable system of equations $f_{j\mu}(\xi) = 0$ has a solution in $\mathfrak{K}$, and hence that every finite subsystem is consistent. By the Hilbert Nullstellensatz in infinite dimensional space [7], it follows that the system also has a solution in $\mathfrak{k}$, and such a solution yields a specialization of (x) into $\mathfrak{o}_1$.

By an analogous argument, we can also prove the following statement: Let $\mathfrak{m}$ be the maximal ideal of $\mathfrak{o}_1$. Let $f(X)$ be a polynomial with coefficients in $\mathfrak{o}_1$. The necessary and sufficient condition that f has a zero in $\mathfrak{o}_1$ is that $f \equiv 0 \pmod{\mathfrak{m}^\nu}$ is solvable in $\mathfrak{o}_1$ for each positive integer ν.

The problem is of course to relax the cardinality assumption, but we see no way of doing this at present.

THE INSTITUTE FOR ADVANCED STUDIES.

REFERENCES.

[1] E. Artin, "*Algebraic Numbers and Algebraic Functions*," Princeton and New York Universities, 1951.

[2] C. Chevalley, "Introduction to the Theory of Algebraic Functions of One Variable," *Mathematical Surveys* No. VI, American Mathematical Society, New York, 1951.

[3] A. Harnack, "Uber die Vieltheiligkeit der ebene algebraische Curve," *Mathematische Annalen*, vol. 10 (1876), pp. 189-198.

[4] Y. Kawada, "Uber die Riemannsche Fläche algebraischer Funktionen," *Proceedings of the Imperial Academy*, Tokyo, vol. 14 (1938), pp. 160-165.

[5] S. Lang, "On quasi algebraic closure," *Annals of Mathematics*, ser. 2, vol. 55 (1952), pp. 373-390.

[6] ———, "The theory of real places," *Annals of Mathematics*, ser. 2, vol. 57 (1953), pp. 378-391.

[7] ———, "Hilbert's Nullstellensatz in infinite dimensional space," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 407-410.

[8] A. Weil, "Foundations of algebraic geometry," *American Mathematical Society Colloquium Publications*, New York, 1946.

[9] E. Witt, "Zyklische Körper und Algebren der Characteristik p vom Grad p^n," *Journal fur Reine und Angewandte Mathematik*, vol. 176 (1937), pp. 126-140.

[10] O. Zariski, "Local uniformization on algebraic varieties," *Annals of Mathematics*, ser. 2, vol. 41 (1940), pp. 852-896.

ON THE EXISTENCE OF CHARACTERS IN TOPOLOGICAL GROUPS.*

By M. COTLAR [1] and R. RICABARRA.

---

1. Introduction. In this section we will explain the idea of the paper and state the main theorems. The proofs will be given in Section 4.

Let $G = \{x, y, \cdots, s, t, \cdots\}$ denote an abelian group, e the unit element, $\chi(x)$ a character of G, $Q = \{f(x)\}$ the set of all uniformly continuous functions on G, and $P = \{\phi(x)\}$ the set of all positive definite functions $\phi(x) \in Q$.

Let us recall that $\phi(x)$ is said to be positive definite if

$$\sum_i \sum_j \lambda_i \bar{\lambda}_j \phi(s_i s_j^{-1}) \geqq 0,$$

for any system of n complex numbers $\lambda_1, \cdots, \lambda_n$ and n elements $s_1, \cdots, s_n \in G$. By a character we mean a continuous function such that

$$|\chi(x)| = 1, \qquad \chi(s^{-1}t) = \overline{\chi(s)}\chi(t).$$

If $\chi(x)$ is not continuous we will say that it is a non-continuous character. Any character, as well as any function of the form $\phi = \sum \lambda_i \chi_i$ with all $\lambda_i \geqq 0$, is positive definite.

G is said to have sufficiently many characters if given an element $s \neq e$ there is a character $\chi(x)$ such that $\chi(s) \neq \chi(e) = 1$.

The aim of this paper is to characterize the class of topological groups with sufficiently many characters. More precisely: a topological group is defined by the group operation xy and the system $\{0\}$ of its open sets; we want to give a condition on the group operation and the system $\{0\}$ which is necessary and sufficient for the existence of sufficiently many characters on G.

Observe that by definition of a topological group, given an integer $n > 0$ and an element $s \neq e$, there is an open set 0 such that $e \in 0$, $s \bar{\in} 0^2 = 0 \cdot 0$. Let us say that a subset $B \subset G$ is a *big set* if G can be covered by a finite number of translations of B: $G \subset s_1 B \cup \cdots \cup s_n B$, $s_i \in G$. For example if G is a compact group then any open set is a big one. We will prove the following lemma.

---

* Received January 23, 1953.

[1] Fellow of the John Simon Guggenheim Foundation.

LEMMA. *In any topological group G, given $s \neq e$, there is a big neighborhood 0 of e such that $e \in 0$ and $s \bar{\in} 0^2 = 0 \cdot 0$.*

We will prove that the lemma is true for 0^n with $n \geqq 6$, if and only if G possesses sufficiently many characters. More precisely:

THEOREM 1. *For any topological group G and any element $s \in G$, $s \neq e$, the following conditions are equivalent:*

(1) *There is a character $\chi(x)$ such that $\chi(s) \neq \chi(e) = 1$.*

(2) *There is a big neighborhood 0 such that $e \in 0$ and $s \bar{\in} 0^6$.*

Evidently the set 0^6 may be replaced by any 0^n, $n \geqq 6$. By the lemma, Theorem 1 is no longer true if 0^6 is replaced by 0^n, $n \leqq 2$. It would be interesting to know if 0^6 can be replaced by 0^n, $n \geqq 3$, or to determine the minimum value for n.

The proof of Theorem 1 is based on another theorem (Theorem 2 below) which we are going to state. We will recall first some known facts from the classical theory. A fundamental result is the following ([1], [2]):

(A) *In any locally compact abelian group there exist sufficiently many characters.*

More precise information is provided by the Tauberian theorems ([3], [4]). Let us say that a function $f(x)$ belongs to the *w^c-closure* of the set of functions $\Lambda = \{g(x)\}$, if given $\epsilon > 0$ and a compact subset $C \subset G$ there is a function $g(x) \in \Lambda$ such that $|f(x) - g(x)| < \epsilon$ for any $x \in C$. On the set of characters $G^* = \{\chi(x)\}$, the w^c-closure defines a topology known as the Pontrjagin topology of G^*. We will denote by $S^c[\Lambda]$ the w^c-closure of the set of all functions of the form $\sum_1^n \lambda_i g_i(s_i x)$, where $s_i \in G$, $g_i \in \Lambda$, and λ_i are complex numbers. Finally we define the *w^c-spectrum* $\sigma^c[\Lambda]$ of the set Λ, as the set of all characters $\chi(x)$ such that $\chi \in S^c[\Lambda]$. If Λ is composed of a single function $f(x)$ then we shall write $S^c[\Lambda] = S^c[f]$, and $\sigma^c[\Lambda] = \sigma^c[f]$ = spectrum of $f(x)$.

The tauberian theorems give us the following properties of the w^c-spectrum of $f(x)$.

(B) *If G is locally compact and $f \in Q$, then the spectrum $\sigma^c[f]$ is a non empty set.*

(B′) *If G is locally compact and $V = \{\chi(x)\}$ is a set of characters such that $\sigma^c[f] \subset V$ and such that V (as a subset of G^*) is open in the w^c-topology, then $f \subset S^c[V]$.*

The hypothesis that V is open in the Pontrjagin topology is essential. In other words, in general the relation $f \subset S^c[\sigma^c[f]]$ is not true. Godement ([5]) proved that this relation is true in the following case:

(C) *Let G be locally compact. If $f \in Q$ is integrable with respect to the Haar measure, or if f is positive definite, then $f \subset S^c[\sigma^c[f]]$.*

It is easy to see that (A) is a corollary of (C) as well as of (B'). The proofs of (A), (B), (C), use Haar measure, the theory of convolution and the theory of Banach Algebras or Hilbert spaces. Nevertheless in the statements of these theorems appear only ordinary notions concerning continuous functions. It is natural therefore to pose the following problem: find a direct proof of (B), (B') and (C), and generalize these theorems for not necessarily locally compact groups. Theorem 2 below gives us a solution for the case of proposition (C). We did not succeed in finding a similar solution for (B').

In order to state Theorem 2 it is necessary to introduce some definitions. Theorem (C) refers to functions integrable with respect to Haar measure and to the w^c-spectrum of f. But neither the Haar measure nor the w^c-topology can be used in general topological groups. Therefore first of all we have to define a type of spectrum and a measure which will fit for any group.

Definition 1. We shall say that a function $f(x), x \in G$, belongs to the *w^u-closure* of the set $\Lambda = \{g(x)\}$, if for any given $\epsilon > 0$, there is a subset $\Lambda' \subset \Lambda$ such that:

(1) all the functions of Λ' are uniformly equi-continuous on G,

(2) given $x_1, \cdots, x_n \in G$ there is a function $g(x) \in \Lambda'$ such that $|f(x_i) - g(x_i)| < \epsilon$ for $i = 1, \cdots, n$.

We shall say that $f(x)$ belongs to the *w-closure* of Λ if only the condition (2) is satisfied.

We shall denote by $S^u[\Lambda]$ the w^u-closure of the set $\{\sum \lambda_i g_i(s_i x)\}$, $g_i \in \Lambda$, of all linear combinations of translations of functions $g_i(x) \in \Lambda$. While $S^c[S^c[\Lambda]] = S^c[\Lambda]$ is true, $S^u[S^u[\Lambda]] = S^u[\Lambda]$ is not generally true. For this reason we define the *σ^u-spectrum* $\sigma^u[\Lambda]$ as the set of all characters $\chi(x)$ such that $\chi(x) \in S^u[\Lambda] \cup S^u[S^u[\Lambda]] \cup S^u[S^u[S^u[\Lambda]]] \cup \cdots$.

If Λ is composed of a single function $f(x)$ then we write $S^u[f]$ and $\sigma^u[f]$ instead of $S^u[\Lambda]$ and $\sigma^u[\Lambda]$.

Definition 2. A function $f \in Q$ will be said *regular* if $f \subset S^u[\sigma^u[f]]$. The set of regular functions is denoted by R.

An immediate consequence of the above definitions is the following proposition.

PROPOSITION. (a) *If* $f_n \in S^u[\Lambda]$, $n = 1, 2, \cdots$, *and* $f_n(x) \to f(x)$ *uniformly on* G, *then* $f(x) \in S^u[\Lambda]$. (b) *If* $f_n \subset S^u[f]$, $n = 1, 2, \cdots$, $f \in Q$, *and each* $f_n(x)$ *is regular, and* $f_n \to f$ *uniformly on* G, *then* $f(x)$ *is also regular.*

Now, if G is not locally compact or locally bounded, it cannot have a Haar measure (that is a measure which is σ-additive, invariant, finite on bounded sets and positive on open sets). But in G there always exist measures $\mu(A)$, $\mu(G) \neq 0$, with the following properties:

(1) μ is *finitely additive*: $\mu(\sum_1^n A_i) = \sum_1^n \mu(A_i)$ for every n disjoint measurable sets.

(2) μ is invariant: $\mu(sA) = \mu(A)$ if $s \in G$, and $\mu(A) = \mu(A^{-1})$.

(3) $\lim_{s \to e} \int |f(sy) - f(y)| \, d\mu(y) = 0$,

for any uniformly continuous and integrable function $f(x)$.[2]

(4) If $f(x) \geqq 0$ is μ-integrable and $\epsilon > 0$, there is a set $C \subset G$ such that $\mu(C) < \infty$ and $\int_{G-C} |f| \, d\mu < \epsilon$.

Such measures μ will be called *G-measures*. We introduce now the three following classes R_0, R_1, R_2 of continuous functions.

Definition 3. Let $f(x) \in Q$. We shall write $f \in R_0$, if

$$f(x) = \int \phi(xy) k(y^{-1}) \, d\mu(y),$$

where $\phi \in P$, μ is a G-measure and $k(y)$ and $|k(y)|$ are μ-integrable functions. We shall write $f \in R_1$ ($f \in R_2$) if there is a G-measure μ and a sequence $\{k_n(x)\}$ such that f is μ-integrable (f is positive definite), each $k_n(x) \geqq 0$ is μ-integrable and $k_n \in P$, and

$$\int f(xy) k_n(y^{-1}) \, d\mu(y) \to f(x)$$

uniformly on G.

THEOREM 2. *Let* G *be an arbitrary abelian topological group. Then any function* $f \in R_0 \cup R_1 \cup R_2$ *is regular, that is:* $R_0 \subset R$, $R_1 \subset R$, $R_2 \subset R$. *Moreover, the spectrum* $\sigma^u[f]$ *of* f *is a* σ*-compact set (in the* w^u*-topology).*

[2] When the integral is extended over the whole space G we will write $\int f \, d\mu$.

It is easy to see that if G is locally compact then Theorem 2 furnishes a more precise form of (C). In fact, we first observe that by a known property of equi-continuous functions on compact sets and by definition of w^u-closure it is clear that

$$S^u[\Lambda] \subset S^c[\Lambda]; \qquad \sigma^u[f] \subset \sigma^c[f];$$
$$S^u[\sigma^u[f]] \subset S^c[\sigma^c[f]].$$

Hence the relation $f \subset S^u[\sigma^u[f]]$ implies the relation $f \subset S^c[\sigma^c[f]]$. Therefore from Theorem 2 it follows that (C) is true for any function $f \subset R_0 \cup R_1 \cup R_2$.

We next observe that if G is a locally compact group, μ its Haar measure, $f(x) \in Q$, and if we take $k_n(x) \geqq 0$ such that $k_n(x) \in P$, $\int k_n(x)d\mu(x) = 1$, and $k_n(x) = 0$ outside of a small neighborhood of e, then

$$\int f(xy)k_n(y^{-1})d\mu(y) \to f(x)$$

uniformly on G. Thus, *in the case of a locally compact group the class R_1 contains all the integrable functions $f \in Q$, and R_2 contains all the positive definite functions.* Hence (C) is a corollary of Theorem 2. Moreover, Theorem 2 gives us a further information about the spectrum of the function.

2. Elementary observations on characters. Let G be an abelian topological group. The simplest transformations of functions $f(x), x \in G$, are the translations: $T_sf(x) = f(s^{-1}x), s \in G$. It is easy to verify that if a bounded function $f \in Q$ satisfies $T_sf(x) = f(x)f(s^{-1})$, for any $s \in G$, then $f(x)$ is a character or $f \equiv 0$, and conversely. If we denote $f(sx)f(s^{-1})$ by $T^sf(x)$ the above condition may be written $T^sf = f$. Hence if Q' is the set of the functions $f \in Q$ such that $|f(x)| \leqq 1$, then each T^s transforms Q' into itself and the characters are the *non zero fixed points* of Q' under all the transformations T^s. Thus the problem of finding characters is related to the problem of finding fixed points but with the essential condition that the fixed point must be $\neq 0$.[3] In order to get the last condition we will consider instead of the whole set Q' the subset $Q' \cap P$. Since the T^s do not transform P into P it will be necessary to take certain linear combinations of the T^s and proceed as follows.

---

[3] If $f \in Q'$, consider all the functions f_i of the form $f_i(x) = T^{s_1}T^{s_2} \cdots T^{s_k}f(x)$, and let C be the set of all the convex combination $\sum \lambda_i f_i$ of such functions. The w-closure $\bar{C}$ of C is a compact (in the w-topology) convex set of equi-continuous functions and T^s transforms $\bar{C}$ into itself. By Tychonoff's fixed-point theorem there is a fixed point in $\bar{C}$. If we could be sure that this point is $\neq 0$, we would obtain the Wiener tauberian theorem (B). Thus Wiener's theorem may be regarded as a kind of a "non-zero" fixed-point theorem.

For any $s \in G$, $0 \leqq \theta \leqq 2\pi$, $a \geqq 0$, we define the following transformations

$$T(a,s,\theta)f(x) = a^2[2f(x) + e^{2i\theta}f(sx) + e^{-2i\theta}f(s^{-1}x)], \tag{1}$$

$$T(s,\theta)f(x) = T(1,s,\theta)f(x) = 2f(x) + e^{2i\theta}f(sx) + e^{-2i\theta}f(s^{-1}x), \tag{1a}$$

so that $T(a,s,\theta)f(x) = a^2T(s,\theta)f(x)$,

$$T(s,\theta)f(e) = 2f(e) + e^{2i\theta}f(s) + e^{-2i\theta}f(s^{-1}), \tag{1b}$$

$$T(s,\theta)(\lambda f)(x) = \lambda T(s,\theta)f(x). \tag{1c}$$

It is easy to verify that $T(a,s,\theta)\phi$ transforms P into P. Moreover, the non zero fixed points are the characters. More precisely:

LEMMA 1. *Let $f(e) = 1$ and $s \in G$. If there exist two values of θ, θ_1 and θ_2, $\theta_1 \neq \theta_2 \pmod{\pi/2}$, and corresponding two real numbers a_1, a_2, such that*

$$f(x) = T(a_1,s,\theta_1)f(x) = T(a_2,s,\theta_2)f(x) \tag{2}$$

for any $x \in G$, then the relations

$$f(xs) = f(x)f(s); \qquad f(xs^{-1}) = f(x)f(s^{-1}) \tag{3}$$

hold for any $x \in G$.

Proof. Suppose that $f(x)$ satisfies the condition $f(x) = T(a,s,\theta)f(x)$. Since $f(e) = 1$, we deduce that $a \neq 0$, and this condition may be written

$$2f(x) + e^{2i\theta}f(sx) + e^{-2i\theta}f(s^{-1}x) = a^{-2}f(x). \tag{2a}$$

For $x = e$, (2a) gives $a^{-2} = 2 + e^{2i\theta}f(s) + e^{-2i\theta}f(s^{-1})$, and replacing this value of a^{-2} in (2a) we obtain

$$[f(xs) - f(x)f(s)]e^{2i\theta} + [f(xs^{-1}) - f(x)f(s^{-1})]e^{-2i\theta} = 0.$$

If this is true for two values of θ ($\not\equiv \bmod \pi/2$), then each bracket of the last expression must be zero and we obtain (3). This proves the lemma.

Now we will drop the condition $f(e) = 1$.

LEMMA 2. *Suppose that $|f(x)| \leqq 1$, and that the following condition is satisfied: For any s, θ, such that $T(s,\theta)f(e) \neq 0$ there is a real number $a \neq 0$ such that $f(x) = T(a,s,\theta)f(x)$. Then $f(x) = f(e)\chi(x)$ where $\chi(x)$ is a (not necessarily continuous) character.*

Proof. Suppose first that for any $s \in G$ there are two values θ_1, θ_2, $\theta_1 \not\equiv \theta_2 \pmod{\pi/2}$, such that $T(s,\theta_i)f(e) \neq 0$, for $i = 1, 2$. Then by hypothesis we we will have $f(x) = T(a_i,s,\theta_i)f(x)$, $i = 1, 2$. From $f(e) = a_i^2T(s,\theta_i)f(e)$

we deduce $f(e) = c \neq 0$, and we may write $f(x)/c = T(a_i, s, \theta_i)(f/c)(x)$, $i = 1, 2$. By Lemma 1 this implies $f(x)/c = \chi(x)$, $f(x) = c\chi(x) = f(e)\chi(x)$ and in this case Lemma 2 is true. Suppose now that for some $s_0 \in G$ it is not possible to find two values of θ ($\not\equiv \bmod \pi/2$) such that

$$T(s_0, \theta)f(e) = 2f(e) + e^{2i\theta}f(s_0) + e^{-2i\theta}f(s_0^{-1}) \neq 0.$$

Then it must be that $f(e) = f(s_0) = f(s_0^{-1}) = 0$. We shall show that this implies $f(x) \equiv 0$. In fact, suppose that on the contrary, for some $s \in G$ we have $f(s) \neq 0$. Then we can find a θ such that $T(s, \theta)f(e) \neq 0$, and by hypothesis we will have $f(x) = a^2T(s, \theta)f(x)$ for some $a \neq 0$. For $x = e$ this gives $f(e) = a^2T(s, \theta)f(e) \neq 0$, and we arrive at a contradiction. Therefore $f(x) \equiv 0$, and we may write $f(x) = f(e) \cdot 1 = f(e)\chi(x)$. This proves Lemma 2.

LEMMA 3. *For any transformation $T(a, s, \theta)$ with $a \neq 0$, there is a transformation $T^*(a, s, \theta)$ of the form*

(4) $T^*(a, s, \theta)f(x)$

$$= (b^2 + b_1^2)f(x) + bb_1e^{2i(\theta+\pi/2)}f(sx) + bb_1e^{-2i(\theta+\pi/2)}f(s^{-1}x)$$

and two numbers $\lambda > 0$, $\mu > 0$, $\lambda + \mu = 1$, such that

(4a) $$f(x) = \lambda T(a, s, \theta)f(x) + \mu T^*(a, s, \theta)f(x), \qquad \lambda + \mu = 1,$$

holds for any $f(x)$. Besides λ may be chosen to be any number which satisfies the inequalities

(4b) $$0 < \lambda < 1, \qquad \lambda \leqq 4^{-1}a^{-2}.$$

For each $T(a, s, \theta)$ there are many such $T^*(a, s, \theta)$ satisfying (4), (4a), (4b), *and any of them will be called an adjoint transformation of $T(a, s, \theta)$, and denoted by T^*.*

Proof. Condition (4a) will certainly be satisfied if $2\lambda a^2 + \mu(b^2 + b_1^2) = 1$ and $\lambda a^2 - \mu bb_1 = 0$. Thus we must have

(4c) $$(b + b_1)^2 = \mu^{-1}, \qquad (b - b_1)^2 = \mu^{-1}[1 - 4\lambda a^2],$$

with $\lambda > 0$, $\mu > 0$, $\lambda + \mu = 1$, and $0 < 4\lambda a^2 < 1$. It is then sufficient to take $0 < \lambda \leqq 4^{-1}a^{-2}$, $\lambda < 1$, $\mu = 1 - \lambda$. This proves Lemma 3.

Remark. $T(a, s, \theta)$ and $T^*(a, s, \theta)$ both transform P into P, for any determination of the adjoint transformation T^*.

3. T-invariant sets. It is well known that a function $f(x)$, $x \in G$, belongs to the w-closure (Definition 1, Section 1) of the set $\Lambda = \{g(x)\}$, if

and only if there is a directed subset $\{g_\alpha(x)\}$ of Λ such that $\lim_\alpha g_\alpha(x) = f(x)$, for any $x \in G$. By a theorem of Tychonoff any w-closed and bounded set $\Lambda = \{g(x)\}$ is w-compact (we say that Λ is bounded if $|g(x)| \leqq M$ for some constant M and any $g \in \Lambda, x \in G$). For such sets we have the following theorem of Minkowski-Krein-Milman: If $\Lambda = \{g(x)\}$ is w-compact and convex (that is $g_i \in \Lambda$, $\lambda_i > 0$, $\sum \lambda_i = 1$, imply $\sum \lambda_i g_i \in \Lambda$), then any $g \in \Lambda$ belongs to the w-closure of the convex combinations of the extreme points of Λ. That is, g belongs to the w-closure of the set of the functions of the form $\sum \lambda_i g_i$, $\lambda_i > 0$, $\sum \lambda_i = 1$, where each g_i is an extreme point of Λ. Let us recall that $g_0 \in \Lambda$, is said to be an extreme point of Λ, if

$$g_0 = \lambda g_1 + \mu g_2, \quad \lambda > 0, \quad \mu > 0, \quad \lambda + \mu = 1, \quad g_1 \in \Lambda, \quad g_2 \in \Lambda,$$

imply $g_0 = g_1 = g_2$.

Definition 4. We shall say that the set $\Lambda = \{f(x)\}$ is *T-invariant* if the following condition is satisfied: If $f \in \Lambda$, and s, θ, are such that $T(s, \theta)f(e) \neq 0$, then there is a number $a \neq 0$, such that $T(a, s, \theta)f(x) \in \Lambda$, and such that *at least for one* determination of T^* we have also $T^*(a, s, \theta)f(x) \in \Lambda$.

Definition 4a. We shall say that $\Lambda = \{f(x)\}$ is *strongly T-invariant* if the following condition is satisfied: if $f \in \Lambda$ and s, θ, are such that $T(s, \theta)f(e) \neq 0$, then there exists a number $a > 0$ such that:

$$0 < m < a^2 \leqq M \cdot (|T(s, \theta)f(e)|)^{-1},$$

(m, M, independent of $f \in \Lambda$ and of s, θ), and

$$T(a, s, \theta)f(x) \in \Lambda, \qquad T^*(a, s, \theta)f(x) \in \Lambda,$$

for any determination of T^*.

Lemma 2 combined with the Krein-Milman theorem yields at once the following method for obtaining characters.

PROPOSITION 1. *Let $\Lambda = \{f(x)\}$ be a set of functions which is convex, w-closed, T-invariant and bounded: $|f(x)| \leqq C$ for all $f \in \Lambda$. Then each extreme point f_0 of Λ is of the form $f_0(x) = f_0(e)\chi(x)$, where $\chi(x)$ is a (not necessarily continuous) character. Hence any $f \in \Lambda$ belongs to the w-closure of the convex combinations $\sum \lambda_i c_i \chi_i(x)$, $\sum \lambda_i = 1$, $\lambda_i \geqq 0$, where each $\chi_i(x)$ is a character and $c_i\chi_i(x) \in \Lambda$. In particular if $f(e) \geqq 0$ for any $f \in \Lambda$, then any $f \in \Lambda$ is positive definite.*

Of course, if Λ is composed of continuous functions then the extreme points of Λ are continuous characters.

Proof. Let f_0 be an extreme point of Λ. If $T(s,\theta)f_0(e) \neq 0$, then by the T-invariance of Λ we have $T(a,s,\theta)f_0(x) \in \Lambda$, $T^*(a,s,\theta)f_0 \in \Lambda$, $a^2 > 0$, and by definition of T^*,

$$f_0 = \lambda T(a,s,\theta)f_0 + \mu T^*(a,s,\theta)f_0, \quad \lambda + \mu = 1, \quad \lambda > 0, \quad \mu > 0.$$

Since f_0 is an extreme point of Λ, this implies $f_0(x) = T(a,s,\theta)f_0(x)$, $a^2 > 0$. Hence by Lemma 2 we have $f_0(x) = f_0(e)\chi(x)$. In particular if $f_0(e) \geqq 0$, then $f_0(x) = f_0(e)\chi(x) \in P$. The rest of the proposition is a consequence of the Tychonoff and Krein-Milman theorems.

The following theorem simplifies the application of Proposition 1.

PROPOSITION 2. *If $\Lambda = \{f(x)\}$ is strongly T-invariant, then the w-closure $\bar{\Lambda}$ of Λ is T-invariant.*

Proof. Let $f \in \bar{\Lambda}$ and $|T(s,\theta)f(e)| > 0$. Then there is a directed set $\{f_\alpha(x)\} \subset \Lambda$ which converges to $f(x) : f_\alpha(x) \to f(x)$ for each $x \in G$, and since $T(s,\theta)f_\alpha(x) \to T(s,\theta)f(x)$, we may suppose that

$$|T(s,\theta)f_\alpha(e)| > \gamma > 0 \text{ for all } f_\alpha. \tag{5}$$

By hypothesis, for any f_α we can find a number $a_\alpha > 0$ such that

$$0 < m < a_\alpha{}^2 \leqq M\gamma^{-1}, \quad T(a_\alpha,s,\theta)f_\alpha \in \Lambda, \quad T^*(a_\alpha,s,\theta)f_\alpha \in \Lambda \tag{6}$$

for any determination of T^*. By Lemma 3 we can take all the $\lambda_\alpha = \lambda$, and all the $\mu_\alpha = \mu$, and

$$f_\alpha = \lambda T(a_\alpha,s,\theta)f_\alpha + \mu T^*(a_\alpha,s,\theta)f_\alpha, \tag{6a}$$

$$\lambda = \delta \cdot 4^{-1}M^{-1} > 0, \quad \mu = 1 - \lambda, \quad 0 < \delta \leqq 1.$$

From (6) it follows that the directed set $\{a_\alpha{}^2\}$, $a_\alpha{}^2 \leqq M : \gamma$, $\gamma > 0$, has a limit point a^2, $0 < m \leqq a^2 \leqq M : \gamma$. For this a^2 the function $T(a,s,\theta)f(x)$ belongs to the w-closure of the set $\{T(a_\alpha,s,\theta)f_\alpha\}$, and from (6a) and (4c) it follows that at least one $T^*(a,s,\theta)f$ belongs to the w-closure of $\{T^*(a_\alpha,s,\theta)f_\alpha\}$. Therefore, from (6) it follows that $T(a,s,\theta)f(x)$ and at least one $T^*(a,s,\theta)f(x)$ belong to $\bar{\Lambda}$, and since $a > 0$, this proves the T-invariance of $\bar{\Lambda}$, and hence Proposition 2 is proved.

Now, let $f(x)$ be a function such that $f(e) = 1$. Consider all the functions $g(x)$ of the form $g(x) = T_1T_2 \cdots T_nf(x)$, $g(e) = 1$, where T_i is any transformation of the form $T(a,s,\theta)$ or $T^*(a,s,\theta)$. *We will denote by $\Delta[f]$* the set of all convex combinations of the form $\sum \lambda_i g_i$, $\lambda_i > 0$, $\sum \lambda_i = 1$, where the g_i are of the above type.

PROPOSITION 3. *A function $\phi(x)$, $\phi(e) = 1$, belongs to P if and only if $\Delta[\phi]$ satisfies the two following conditions:*

(1) $|f(x)| \leqq 1$ *for any* $f \varepsilon \Delta[\phi]$;

(2) $T(s, \theta)f(e) \geqq 0$ *for any* s, θ, *and* $f \varepsilon \Delta[\phi]$.

Furthermore, in this case $\Delta[\phi]$ is strongly T-invariant with $m = \frac{1}{4}$, $M = 1$.

Proof. If $f \varepsilon P$ then also $T(a, s, \theta)f \varepsilon P$, $T^*(a, s, \theta)f \varepsilon P$, $|f(x)| \leqq f(e) \geqq 0$. Hence if $\phi \varepsilon P$ then $f \varepsilon P$, for any $f \varepsilon \Delta[\phi]$, and $f(e) = 1$ by definition of $\Delta[\phi]$. Therefore $|f(x)| \leqq f(e) = 1$, and $T(s, \theta)f(e) \geqq 0$. Conversely, suppose that any $f \varepsilon \Delta[\phi]$ satisfies (1) and (2). If $T(s, \theta)f(e) = c \neq 0$, then by (2) $c > 0$, hence if $a^2 = c^{-1}$ then $T(a, s, \theta)f(e) = 1$, and therefore $T(a, s, \theta)f(x) \varepsilon \Delta[\phi]$. From $f(e) = T(a, s, \theta)f(e) = 1$, and from

$$f(e) = \lambda T(a, s, \theta)f(e) + \mu T^*(a, s, \theta)f(e)$$

it follows that $T^*(a, s, \theta)f(e) = 1$, and consequently $T^*(a, s, \theta)f \varepsilon \Delta[\phi]$. Since $|f(x)| \leqq 1$ and $a^{-2} = T(s, \theta)f(e) = 2f(e) + e^{2i\theta}f(s) + e^{-2i\theta}f(s^{-1}) \leqq 4$, it follows that $\Delta[\phi]$ is strongly T-invariant with $m = \frac{1}{4}$, $M = 1$. By Proposition 2, $\bar{\Delta} = \overline{\Delta[\phi]}$ is T-invariant, and being w-closed, convex and bounded, by Proposition 1, ϕ belongs to the w-closure of the functions $\sum \lambda_i c_i \chi_i$, $\sum \lambda_i = 1$, $\lambda_i \geqq 0$, where $c_i \chi_i = f_i \varepsilon \Delta[\phi]$ and $\chi_i =$ character. Since $c_i = f_i(e) = 1$, all the functions $\sum \lambda_i c_i \chi_i$ are positive definite and therefore $\phi \varepsilon P$.

Remark. The same argument proves that condition (1) alone is necessary and sufficient for ϕ to be a linear combination of positive definite functions. An extreme case of condition (1) is when $\Delta[\phi]$ is formed by a single function. Thus, if $\Delta[\phi]$ is contained in a "circle" of "centre" ϕ and "radius $r < \infty$," then ϕ is a linear combination of positive definite functions; if the "radius" r is 0, then ϕ is a character.

COROLLARY 1. *If $\phi \varepsilon P$, $\phi(e) = 1$, then each extreme point of $\bar{\Delta} = \overline{\Delta[\phi]}$ is a (not necessarily continuous) character.*

It is important to observe that even if ϕ is continuous and positive definite, $\phi(e) = 1$, we cannot be sure that the extreme points of $\overline{\Delta[\phi]}$ are continuous functions. This difficulty will be avoided in the next section by submitting the set $\Delta[\phi]$ to a convolution with a continuous kernel.

4. Proof of Theorems 1 and 2. In this section we prove the results stated in Section 1.

Proof of the Lemma. Let $s \neq e$, and take an open set U such that

$e \subset U$, $s \bar{\varepsilon} U^2$, $U = U^{-1}$. Let $\{B\}$ be the class of all sets B with the following property: $U \subset B$, $B = B^{-1}$, $s \bar{\varepsilon} B^2$. It is easy to see that Zorn's lemma is applicable here so that there is a set $0 \varepsilon \{B\}$ which is maximal with respect to the above property, in other words: $B \varepsilon \{B\}$, $0 \subset B$ imply $0 = B$. We shall show that 0 is a big set.

Let $x \varepsilon G$. If $x \bar{\varepsilon} 0$, the set $0 \cup x \cup x^{-1}$ cannot belong to $\{B\}$, because of the maximal property of 0, hence we must have

$$s \varepsilon (0 \cup x \cup x^{-1})^2 = 0^2 \cup x0 \cup x^{-1}0 \cup x^2 \cup x^{-2}.$$

Since $s \bar{\varepsilon} 0^2$ we have only the following four possibilities: $s \varepsilon x0$, or $s \varepsilon x^{-1}0$, or $s = x^2$, or $s = x^{-2}$. Hence for any element $x \varepsilon G$ we have only three possibilities: $x \varepsilon 0 \cup s0 \cup s^{-1}0$, or $x^2 = s$, or $x^2 = s^{-1}$. Let C be the set of all $x \varepsilon G$ such that $x^2 = s$, and C' the set of x such that $x^2 = s^{-1}$. Then $G \subset \{0 \cup s0 \cup s^{-1}0\} \cup C \cup C'$.

Take a fixed element $x_0 \varepsilon C$ and let $C_0 = x_0^{-1}C$. Then if $y \varepsilon C_0$ it follows that $y^2 = e$. If $y \varepsilon C_0$, $y \bar{\varepsilon} 0$, then as above we will see that $s \varepsilon (0 \cup y \cup y^{-1})^2 = 0^2 \cup y0 \cup y^{-1}0$. Hence $y \varepsilon s0$ or $y \varepsilon s^{-1}0$. Thus $C_0 \subset 0 \cup s0 \cup s^{-1}0$ and $C \subset x_0 0 \cup x_0 s0 \cup x_0 s^{-1}0$. Similarly

$$C' \subset x_0' 0 \cup x_0' s0 \cup x_0' s^{-1}0, \ .$$

and we obtain

$$G \subset \{0 \cup s0 \cup s^{-1}0\} \cup \{x_0 0 \cup x_0 s^{-1}0 \cup x_0 s0\} \cup \{x_0' 0 \cup x_0' s0 \cup x_0 s^{-1}0\}.$$

Hence G is covered by nine translations of 0, hence 0 is a big set. Besides, $e \subset U \subset 0$ is interior to 0 and $s \bar{\varepsilon} 0^2$. This proves the lemma.

Now we will prove Theorem 1 in the following more complete form.

THEOREM 1 bis. *Let G be an abelian topological group and $s \varepsilon G$, $s \neq e$. Then the following three conditions are equivalent:*

(a) *There is a character $\chi(x)$ such that $\chi(s) \neq \chi(e) = 1$.*

(b) *There is a big open set 0 such that $e \varepsilon 0$, $s \bar{\varepsilon} 0^6$, $0 = 0^{-1}$.*

(c) *There is a function $f \varepsilon R_0$ such that $f(s) \neq f(e)$.*

Proof. (a) *implies* (b): Since $\chi(s) \neq 1$, $|\chi(s)| = 1$, we can find an arc l of the unit circle $|z| = 1$ such that $1 \varepsilon l$ and $\chi(s) \bar{\varepsilon} l^6$, and l is symmetric with respect to 1. It is easy to see, since χ is a continuous homomorphism of G into the unit circle, that $0 = \chi^{-1}(l)$ satisfies condition (b). Moreover, the same argument shows that (a) implies (b) with any exponent 0^n, $n \geqq 1$.

(b) *implies* (c): It was proved by Banach and v. Neumann that there exists a finitely additive measure $\mu(A)$ defined on all the subsets $A \subset G$ (so that there is no question about measurability), and such that: $\mu(G) = 1$, $\mu(A) \geqq 0$, $\mu(sA) = \mu(A)$ for any $s \in G$.[4] μ may be taken to satisfy also $\mu(A) = \mu(A^{-1})$ because the measure $2^{-1}\{\mu(A) + \mu(A^{-1})\}$ has this property.

Let 0 be a big open set such that $e \in 0$, $s \bar{\in} 0^6$ and define

$$\phi(x) = \{\mu(0^2)\}^{-1} \int f_{0^2}(xy) f_{0^2}(y^{-1}) d\mu(y),$$

where f_A denotes the characteristic function of the set A. From the invariance of μ, it follows easily that ϕ is a positive definite function (this is very well known if μ is Haar measure, but the proof does not use the σ-additivity of the measure and applies to any finitely additive invariant measure).

Let $k(x)$ be a uniformly continuous function equal to 1 if $x \in 0$ and to zero if $x \bar{\in} 0^2$, and $0 \leqq k(x) \leqq 1$. Such a function exists by a well known argument of Urysohn adapted by Kakutani for topological groups. Since $\mu(G) = 1$ and all the bounded functions are μ-integrable, the function f defined by $f(x) = \int k(xy)\phi(y) d\mu(y)$ belongs to the class R_0. We shall prove that $f(s) \neq f(e)$.

Since 0 is a big set, a finite number of translations of 0 covers G, hence by the invariance of μ we must have $\mu(0) > 0$, $\mu(0^2) > 0$. From the definition of ϕ it follows that $\phi(x) \geqq \mu(0) : \mu(0^2)$ if $x \in 0$, and $\phi(x) = 0$ if $x \in G - 0^4$. Therefore, since $k(x) = 0$ for $x \in G - 0^2$, $f(x)$ vanishes outside of 0^6. In particular $f(s) = 0$. But for $x \in 0$ we have $k(x) = 1$, $\phi(x) \geqq \mu(0) : \mu(0^2)$, hence

$$f(e) = \int k(y)\phi(y) d\mu(y) \geqq \{\mu(0)\}^2 : \mu(0^2) > 0.$$

Thus, $f \in R_0$ and $f(s) \neq f(e)$.

(c) *implies* (a): This follows immediately from Theorem 2. In fact by this theorem $f \in R_0$ may be uniformly approximated on any finite set by a function of the form $\sum \lambda_i \chi_i(x)$. Since $f(s) \neq f(e)$, one at least of the characters χ_i must satisfy $\chi_i(s) \neq \chi_i(e)$, and this proves Theorem 1.

Thus, it only remains to prove Theorem 2.

Proof of Theorem 2. (1) $R_0 \subset R$: Let

$$f(x) = \int k(\psi y^{-1}) \phi(y^{-1}) d\mu(y) = \int \phi(xy) k(y^{-1}) d\mu(y)$$

be a function belonging to the class R_0, where $\phi \in P$, $\phi(e) = 1$, and $k(y)$

[4] A more general theorem is given in Banach's book [6], page 28.

is μ-integrable and $k(y) \geqq 0$. Consider the set $\Delta[\phi] = \{h(x)\}$ and for each $h \in \Delta[\phi]$ define

$$(7) \quad g(x) = k * h(x) = \int k(xy)h(y^{-1})d\mu(y) = \int h(xy)k(y^{-1})d\mu(y),$$

and let $K = \{g(x) = k * h(x)\}$. Since each $h \in \Delta[\phi]$ is positive definite we have $|h(x)| \leqq 1$, $|T(a,s,\theta)h(x)| \leqq T(a,s,\theta)h(e)$, and since

$$(7a) \qquad T(a,s,\theta)g(x) = \int T(a,s,\theta)h(xy)k(y^{-1})d\mu(y)$$

we obtain

$$(8) \qquad |g(x)| \leqq \int |k(y)|\, d\mu(y) = C,$$

$$(8a) \qquad |T(a,s,\theta)g(e)| \leqq |T(a,s,\theta)h(e)| \cdot C.$$

By Proposition 3, $T(s,\theta)h(e) \neq 0$ implies $T(a,s,\theta)h(x) \in \Delta[\phi]$ with $\frac{1}{4} \leqq a^2 \leqq \{T(s,\theta)h(e)\}^{-1}$. Hence from (8a) we derive that K is strongly T-invariant with $m = \frac{1}{4}$, $M = C$. Therefore $\bar{K} = w$-closure of K is T-invariant (Proposition 2), and being convex and bounded (by (8), it follows from Proposition 1 that $f \in K \subset \bar{K}$ belongs to the w-closure of the functions $\sum \lambda_i c_i \chi_i(x)$ where $\lambda_i > 0$, $\sum \lambda_i = 1$, $c_i\chi_i(x) \in \bar{K}$, $\chi_i =$ character. But now we can see that these characters χ_i *are continuous functions,* because from $|g(tx) - g(x)| \leqq \int |k(ty) - k(y)|\, d\mu(y)$ it follows that the functions $g \in K$ are uniformly equi-continuous on G, and hence the limit functions $g \in \bar{K}$ are also uniformly equi-continuous. Furthermore, on account of this equi-continuity the w-closure of K becomes the w^u-closure. From the definition of $f(x)$ and $g \in K$ it is clear that $g \in S^u[f]$, hence $\bar{K} \subset S^u[f]$ and all the above characters χ_i belong to $S^u[f]$, hence to $\sigma^u[f]$. Thus f belongs to the w-closure of the functions $\sum \lambda_i c_i \chi_i(x)$, $\chi_i \in \sigma^u[f]$, and by the equi-continuity it follows that $f \in S^u[\sigma^u[f]]$. Hence $f \in R$.

(2) $R_1 \subset R$: Let $f \in R_1$, so that

$$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \int f(xy)k_n(y)d\mu(y) = \lim_{n\to\infty} \int f(y)k_n(xy)d\mu(y),$$

uniformly on G, and f is μ-integrable, $k_n \in P$ and is μ-integrable. By the part (1) of the theorem, $f_n \in R$ for each n. Let us fix n and consider the set of all functions of the form $\sum \lambda_i f(xy_i)k_n(y_i)$ where

$$\sum |\lambda_i k_n(y_i)| \leqq \int k_n(y)\, d\mu(y) + \epsilon.$$

This is a set of equi-continuous functions, for

$$|\sum \lambda_i f(xy_i)k_n(y_i) - \sum \lambda_i f(x'y_i)k_n(y_i)|$$
$$\leqq \sup_i |f(xy_i) - f(x'y_i)| \{\int k_n(y)\, d\mu + \epsilon\}.$$

From the definition of the integral and the property (4) of G-measures it follows easily that f_n belongs to the w^c-closure of this set, hence by the equi-continuity also to the w^u-closure. Hence $f_n \in S^u[f]$. Since $f_n \in R$, $f_n \to f$ and $f_n \in S^u[f]$, it follows from the Proposition of Section 1 that $f \in R$. Similarly, it is proved that $R_2 \subset R$, and this proves Theorem 2.

Remark. Let R_3 be the set of functions $f \in Q$ such that

$$f(x) = \lim_{n\to\infty} \int f(xy) k_n(y) d\mu(y)$$

uniformly on G, where $k_n \geqq 0$ belongs to P and is μ-integrable, and f is a linear combination of positive definite functions. Modifying the above proof we can prove that $R_3 \subset R$.

UNIVERSITY OF CHICAGO,
UNIVERSITY OF BUENOS AIRES.

BIBLIOGRAPHY.

[1] D. Raikov, "Harmonic analysis on commutative groups with the Haar measure and the theory of characters," *Travaux de l'Intitut Mathématique Stekloff*, vol. 14 (1945).

[2] H. Cartan and R. Godement, "Théorie de la dualité et analyse harmonique dans les groupes abéliens localement compact," *Annales de l'École Normale Supérieure*, vol. 64 (1947).

[3] I. Segal, "The group algebra of a locally compact group," *Transactions of the American Mathematical Society*, vol. 61 (1947).

[4] R. Godement, "Théorèmes taubériens et théorie spectrale," *Annales de l'École Normale Supérieure*, vol. 64 (1947).

[5] R. Godement, "Les fonctions de type positif et théorie des groupes," *Transactions of the American Mathematical Society*, vol. 63 (1948).

[6] S. Banach, *Théorie des Opérations Linéaires*, Warszawa (1932).

ON THE TOTAL RELATIVE STRENGTH OF SOME HAUSDORFF METHODS EQUIVALENT TO IDENTITY.*

By S. K. Basu.

1. Introduction. Given a sequence $\{s_n\}$ and a number α (the numbers with which we are concerned here are all supposed to be real, unless the contrary is stated expressly), let S_α denote the matrix of the transformation

$$\sigma_n = \alpha s_n + (1-\alpha)(s_0 + s_1 + \cdots + s_n)/(n+1) \qquad (n = 0, 1, 2, \cdots) \tag{1}$$

so that

$$S_\alpha = \alpha E + (1-\alpha)M, \tag{2}$$

where E is the identity matrix and M the arithmetic mean matrix (for definitions of E and M see [6], p. 105). It has been proved by Mercer that for $\alpha > 0$, S_α (or the method defined by S_α) is equivalent to identity ([11], pp. 207-212—for $\alpha < 0$, this equivalence no longer holds); in symbol $S_\alpha \approx E$, i.e. in (1), $s_n \to l$ implies $\sigma_n \to l$ and conversely (unless otherwise stated, l is finite). Since then different proofs have been given of this important limit theorem by different writers and various generalizations (for reference see [16], p. 14; also see [10]) have been made. However, the case of infinite limits in the above theorem (i.e. when $l = +\infty$) seems to have been discussed only by Mercer himself. It is natural to ask as to what happens when one replaces M in (2) by some suitable regular matrix. We shall be concerned here with a special case of this general question, when M in (2) is replaced by a Hausdorff matrix (H-matrix) of a special kind (for an H-matrix and its properties see [6], Chap. XI; note that M itself is an H-matrix), taking into account the possibility of both finite and infinite limits, the case of finite limits being discussed in § 3 (where our results are obtained mainly by a direct appeal to Pitt's General Mercerian Theorem; see [12], p. 517) and that of infinite limits in § 4. In § 5 we consider the question of total relative strength of two matrices $\bigcup_\alpha$ and $\bigcup_\beta$, $0 < \beta < \alpha$, where

$$\bigcup_\alpha = \alpha E + (1-\alpha)H,$$

* Received July 11, 1953; revised November 9, 1953.

H being a suitable H-matrix and $\bigcup_\beta$ is similarly defined. (For the notion of total relative strength see [4], p. 51. When a method of summability A is stronger than another B, we use the symbol $A \supset B$ and when this is not true we write $A \not\supset B$. In case A is totally stronger than B, we express it by A t.s. B. The symbol A n.t.s. B will denote that A is stronger but not totally stronger than B) The author wishes to consider similar problems for continuous Hausdorff transformations in a subsequent note.

2. The Hausdorff transformation of a sequence $\{s_n\}$ is defined by

$$\sigma_n = \sum_{k=0}^{n} C_k{}^n \Delta^{n-k}\mu_k \cdot s_k \qquad (n = 0, 1, 2, \cdots), \tag{3}$$

where $\Delta^0\mu_k = \mu_k$, $\Delta^n\mu_k = \Delta^{n-1}\mu_k - \Delta^{n-1}\mu_{k+1}$ $(n \geqq 1)$, and $\{\mu_n\}$ is a given *factor sequence.* In case (3) is a regular transformation $\{\mu_n\}$, called a *regular moment sequence,* has the representation

$$\mu_n = \int_0^1 u^n d\chi(u) \qquad (n = 0, 1, 2, \cdots),$$

where $\chi(u)$, termed the *regular mass function,* is a function of bounded variation in the closed interval $< 0, 1 >$. If H denotes a regular H-method defined by $\{\mu_n\}$, we use the symbol $H \sim \mu_n$. To give prominence to the mass function $\chi(u)$ defining H sometimes we use the symbol $H \sim \chi$ (these symbols are due to Rogosinski; see [13], pp. 166-167). The transformation (3), when regular, is said to be *totally regular* if $s_n \to \infty$ implies $\sigma_n \to \infty$. (For conditions of total regularity of the transformation (3) see [3], Lemma 1. The necessary and sufficient condition of total regularity of the transformation (3), in terms of the mass function, is that $\chi(x) \geqq 0$ and $\uparrow$ in $< 0, 1 >$; see [3], proof of Lemma 1. Also cf. [9], Theorem VI and [2], p. 13. Such a mass function will be called a *totally regular mass function*).

By taking different suitable moment functions (for the definition and properties of moment functions see [7], pp. 95, 102-103) we can have different regular H-matrices. Thus $\mu(1) = 1$ $(t \geqq 0)$ gives the identity matrix E. For the matrix M, $\mu(t) = 1/(t+1)$; for H^k, the Hölder matrix of order $k > -1$, $\mu(t) = 1/(t+1)^k$; and for the Euler's matrix E_α of order $\alpha \geqq 1$ (for the Euler's matrix we have used here the notation as in [3]), $\mu(t) = \alpha^{-t}$ (for the corresponding mass functions see [13], p. 167). These matrices are all totally regular.

3. Suppose now that M in (2) is replaced by a totally regular H-matrix $H \sim \mu_n$ corresponding to the mass function $\phi(t)$ which is of regular bounded

variation (r. b. v.) in $< 0, 1 >$ (for the definition of a function of r. b. v. see [13], p. 176). We then consider the method $\bigcup_\alpha$ defined by

$$\bigcup_\alpha = \alpha E + (1-\alpha)H. \tag{4}$$

The following theorem will be proved.

THEOREM 1. (i) *For* $\alpha > \frac{1}{2}$, $\bigcup_\alpha \approx E$.

(ii) *For* $0 < \alpha \leqq \frac{1}{2}$, $\bigcup_\alpha \approx E$,

provided that the moment function $\mu(t)$ *corresponding to* H *satisfies the additional condition*:

(A) *When considered as a function of a complex variable* $z = x + iy$,

$$\Re\mu(z) \geqq 0 \text{ for } x \geqq 0.^1$$

(iii) *If* $\phi(t)$ *is continuous at* $t = 1$, *then for* $\alpha < 0$, $\bigcup_\alpha \supset E$ (*in fact*, $\bigcup_\alpha \supset E$ *irrespective of the continuity of* $\phi(t)$ *at* $t = 1$) *but* $E \not\supset \bigcup_\alpha$.

We require the following lemmas.

LEMMA 1. *Two regular H-methods with non-vanishing moment sequences cannot be totally stronger than each other unless they are identical* (See [3], Lemma 2).

LEMMA 2. *If* $\mu(z)$ *is the moment function of a regular H-matrix* $H \sim \chi$ *such that* $\chi(t)$ *is of* r. b. v. *in* $< 0, 1 >$ *and*

$$|\mu(z)| \geqq d > 0 \qquad (x \geqq 0),$$

then $E \supset H$. (See [12], p. 517 or [13], p. 178).

LEMMA 3. *With the notations of Lemma 2,*

$$\lim_{t\to\infty} \mu(t) = \chi(1) - \chi(1-0).$$

(This is a known result. See, for instance [13], p. 175)

Proof of Theorem 1. Let $\mu(t)$ and $a(t)$ respectively be the moment functions of H and the regular matrix $\bigcup_\alpha \sim a_n$ and let $\chi(t)$ be the mass function of the latter.

---

[1] We have preferred this restriction for simplicity's sake. We might have imposed a less restrictive condition on $\mu(t)$. It may be noted here that in order to extend the result (i) to the case $0 < \alpha \leqq \frac{1}{2}$, some suitable additional condition on $\mu(t)$ is essential. See [8], p. 906, Theorem 5 and the proof thereof. Also cf. [13], p. 188, § 5. 1.

Then $\chi(t)$ is of r. b. v. in $<0,1>$ and $\chi(t)=(1-\alpha)\phi(t)$ $(0\leqq t<1)$, $\chi(1)=1$. The transformation defined by (4) is

$$\sigma_n=\alpha s_n+(1-\alpha)\sum_{k=0}^{n}C_k^n\Delta^{n-k}\mu_k\cdot s_k \qquad (n=0,1,2,\cdots)$$

for which

$$a_n=\alpha+(1-\alpha)\mu_n. \tag{5}$$

Since H is totally regular, $\Delta^{n-k}\mu_k\geqq 0$ $(0\leqq k\leqq n)$. Hence

$$\sum_{k=0}^{n}C_k^n\,|\,\Delta^{n-k}\mu_k\,|=\sum_{k=0}^{n}C_k^n\Delta^{n-k}\mu_k=1.$$

Consequently, $\bigcup_a\approx E$ for $\alpha>\frac{1}{2}$. (See [10], Theorem 1 and Cor. 3. We have deduced (i) directly from some results in [10]. We could have made use of Lemma 2 as well for the purpose. The case for $\alpha>\frac{1}{2}$ also follows easily from [8], Theorem 5 or from [1], Theorems 3. 2, 4. 1. It is easy to give an independent proof of (i)). This proves (i).

To prove (ii) we consider the moment function $a(t)=\alpha+(1-\alpha)\mu(t)$. We have

$$a(x+iy)=\alpha+(1-\alpha)\mu(x+iy)=\alpha+(1-\alpha)(u+iv),\ \text{say},$$

where, by hypothesis, $u=u(x,y)\geqq 0$ for $x\geqq 0$. Then

$$|\,a(x+iy)\,|=|\,\{\alpha+(1-\alpha)u\}+i(1-\alpha)v\,|.$$

Hence, when $0<\alpha<\frac{1}{2}$,

$$|\,a(x+iy)\,|\geqq\alpha+(1-\alpha)u\geqq\alpha>0 \text{ for } x\geqq 0.$$

Therefore, by Lemma 2, $\bigcup_a\approx E$ for $0<\alpha\leqq\frac{1}{2}$.

In order to prove (iii) put $\alpha=-\beta$ $(\beta>0)$. Then by Lemma 3, since $\phi(t)$ is now continuous at $t=1$,

$$a(t)\to\chi(1)-\chi(1-0)=1-(1+\beta)\phi(1)=-\beta. \tag{6}$$

Hence there is a number $\tau>0$, such that $a(t)<0$ for $t\geqq\tau$. Since $a(t)$ is continuous for $t\geqq 0$ with $a(0)=1$, it follows that $a(t_0)=0$ for some $t_0>0$.

Consequently, $E\,\mathfrak{P}\bigcup_a$ (See [13], Remark (iii) on p. 178).

Two particular cases of Theorem (i).

(i) Put $\qquad H=c_1H^1+c_2H^2+\cdots+c_pH^p,$

where $c_r \geqq 0$ $(r = 1, 2, 3, \cdots, p)$ and $\sum_{r=1}^{n} c_r = 1$. Then

$$\alpha E + (1 - \alpha)(c_1 H^1 + c_2 H^2 + \cdots + c_p H^p) \approx E.$$

The corresponding result for continuous Hausdorff transformations follows easily from a general result due to Rogosinski (See [14], p. 359, Theorem 12. The definition of equivalence as used by Rogosinski is, however, somewhat different from ours).

(ii) It is easily seen that, p being a positive integer,

$$S_\alpha{}^p = [\alpha E + (1 - \alpha) M]^p$$

is totally regular for $0 < \alpha < 1$. For, $S_\alpha{}^p$ has the form

$$S_\alpha{}^p = c_1 E + c_2 M + \cdots + c_p M^p,$$

where $c_r \geqq 0$ $(r = 1, 2, 3, \cdots, p)$, and $\sum_{r=1}^{p} c_r = 1$. Hence $\alpha E + (1 - \alpha) S_\alpha{}^p \approx E$.

The corresponding result for continuous Hausdorff transformations also follows easily from a more general result due to Fuchs and Rogosinski (See [5], p. 209, Theorem 5).

In what follows when the index of summability (this nomenclature is due to Rogosinski and Fuchs; see [5], p. 29) for the matrix $\bigcup_\alpha$ is such that $0 < \alpha \leqq \frac{1}{2}$, the moment function $\mu(t)$ is supposed to have satisfied the additional condition (A) stated in Theorem 1. Such a moment function we shall call where necessary a *restricted moment function,* and $\bigcup_\alpha$ corresponding to a restricted moment function will be called a *restricted matrix.*

4. We shall next prove the following Theorem.

THEOREM 2. (i) *For* $0 < \alpha < 1$, $\bigcup_\alpha$ t.s. E *but* E n.t.s. $\bigcup_\alpha$.

(ii) *If* $\alpha > 1$, E t.s. $\bigcup_\alpha$ *but* $\bigcup_\alpha$ n.t.s. E.

(iii) *For* $\alpha < 0$, $\phi(t)$ *being continuous at* $t = 1$, $\bigcup_\alpha$ n.t.s. E.

We require the following lemma.

LEMMA 4. $u(t)$ *being totally monotone for* $t \geqq 0$, $u^n(t)$ $(n = 1, 2, 3, \cdots)$ *are also totally monotone.*

Proof. If $u(t)$ $(=u)$ and $v(t)$ are totally monotone for $t \geqq 0$, *then* by Leibniz Theorem on successsive differentiation

$$(uv)_n = \sum_{r=0}^{n} c_r{}^n u_{n-r} v_r$$

or

$$(-1)^n (uv)_n = \sum_{r=0}^{n} c_r{}^n (-1)^{n-r} u_{n-r} \cdot (-1)^r v_r,$$

so that uv is also totally monotone for $t \geqq 0$. In particular, $\{u(t)\}^2 = u^2(t)$ is totally monotone for $t \geqq 0$. The total montony of $u^n(t)$ then follows by induction.

Proof of Theorem 2. To prove (i) we notice from (4) that, for $0 < \alpha < 1$, $\bigcup_\alpha$ is the sum of two totally regular matrices with positive coefficients and as such is totally regular, as it is regular. Clearly, the moment sequence $\{a_n\}$ as given by (5) is non-vanishing and consequently, by Lemma 1, $\bigcup_\alpha{}^{-1}$ cannot be totally regular.

For the proof of (ii) we observe that, H being totally regular, $\phi(t)$ is non-decreasing in $<0, 1>$. Hence for $t \geqq 0$,

$$0 \leqq \mu(t) = \int_0^1 u^t d\phi(u) \leqq \int_0^1 d\phi(u) = 1.$$

Accordingly, $b(t)$ being the moment function for $\bigcup_\alpha{}^{-1}$, as $\alpha > 1$, $b(t) = 1/a(t)$ $= \sum_{n=0}^{\infty} v_n(t)$, where $v_n(t) = 1/\alpha \cdot (1 - 1/\alpha)^n \mu^n(t)$ $(n = 0, 1, 2, \cdots)$. Also for $t > 0$, provided term-by-term differentiation is permissible,

$$(7) \qquad (-1)^k b^{(k)}(t) = \sum_{n=0}^{\infty} (-1)^k v_n{}^{(k)}(t) \geqq 0 \qquad (k = 0, 1, 2, \cdots),$$

as $v_n(t)$ $(n = 1, 2, 3, \cdots)$ are totally monotone by Lemma 4.

We have still to justify the term-by-term differentiation in (7). If we consider $\mu(t)$ as a function of a complex variable then $\mu(t)$ is regular for $x > 0$ and, therefore, $v_n(t)$ $(n = 0, 1, 2, \cdots)$ are also regular for $x > 0$.

Again, for $x > 0$,

$$|\mu(t)| \leqq \int_0^1 |u^{x+iy}| \, d\phi(u) = \int_0^1 u^x d\phi(u) \leqq \int_0^1 d\phi(u) = 1.$$

Hence $|v_n(t)| < (1 - 1/\alpha)^n$ $(n = 0, 1, 2, \cdots)$. Therefore, by Weierstrass' M-test the series $\sum v_n(t)$ is uniformly convergent for $\Re(t) > 0$ and as such

can be differentiated term-by-term as often as we please (see, for instance, [15], p. 95). The term-by-term differentiation in (7) is thus justified. Also since $b(0) = b(+0)$, it follows from (7) that $b(t)$ is totally monotone for $t \geqq 0$. (The total monotony of $b(t)$ also follows easily from a known Theorem. See [17], p. 151, Cor. 7.) Hence $\bigcup_\alpha^{-1}$ is totally regular and so $\bigcup_\alpha$ cannot be totally regular. The latter result also follows from the fact that, for $\alpha > 1$, $\chi(t) = -(\alpha-1)\phi(t) \leqq 0$ $(0 < t < 1)$, so that $\chi(x) < 0$ for some t and since $\chi(0) = 0$, $\chi(t)$ cannot be non-decreasing in $<0, 1>$. The proof of (ii) is now complete.

In (6) we have already noticed that $\chi(1) - \chi(1-0) = -\beta < 0$; so that in a left-hand neighborhood of $t = 1$, $\chi(t)$ is not a non-decreasing function. Consequently, $\bigcup_\alpha$ cannot be totally regular and (iii) follows.

5. Let $\bigcup_\beta = \beta E + (1-\beta)H$, so that $\bigcup_\beta$ is a matrix obtained from (4) by replacing the index of summability α by β, the matrix H remaining the same. Then $\bigcup_\alpha \approx E \approx \bigcup_\beta$ for $\alpha, \beta > 0$. (In such cases of comparison of the strength or total strength of two matrices $\bigcup_\alpha$, $\bigcup_\beta$, since the same matrix H is to serve in both cases, whenever one of the matrices is restricted, the other is also supposed to be so.) Our object is to compare the total relative strength of the matrices $\bigcup_\alpha$ and $\bigcup_\beta$ for $0 < \beta < \alpha$. In this connection the folloiwng theorems will be proved.

THEOREM 3. *For* $1 < \beta < \alpha$, $\bigcup_\beta$ t.s. $\bigcup_\alpha$ *but* $\bigcup_\alpha$ n.t.s. $\bigcup_\beta$.

THEOREM 4. *If* $0 \leqq \beta < 1 < \alpha$, $\bigcup_\beta$ t.s. $\bigcup_\alpha$ *but for* $0 < \beta < 1 < \alpha$, $\bigcup_\alpha$ n.t.s. $\bigcup_\beta$.

THEOREM 5. *For* $0 < \beta < \alpha < 1$, $\bigcup_\alpha$ n.t.s. $\bigcup_\beta$.

THEOREM 6. *If* $0 \leqq \beta < \alpha < 1$, *there are some* $H \sim \mu_n$ *for which* $\bigcup_\beta$ t.s. $\bigcup_\alpha$, *and other* H *for which* $\bigcup_\beta$ n.t.s. $\bigcup_\alpha$.

Proof of Theorems 3, 4, 5. Let $\{b_n\}$ and $\{M_n\}$ respectively be the moment sequences of $\bigcup_\beta$ and $\bigcup_\beta/\bigcup_\alpha$. Then

$$M_n = b_n/a_n = (\beta-1)/(\alpha-1) + (\alpha-\beta)/(\alpha-1) \cdot 1/a_n.$$

Now, for $\alpha > 1$, $\{1/a_n\}$ is totally monotone by Theorem 2(ii). Hence for $1 < \beta < \alpha$, M_n is totally monotone; consequently, $\bigcup_\beta$ t.s. $\bigcup_\alpha$ and so $\bigcup_\alpha$ n.t.s. $\bigcup_\beta$. This proves Theorem 3.

For $0 \leqq \beta < 1 < \alpha$,

$$M_k = (\alpha-\beta)/(\alpha-1) \cdot 1/a_k - (1-\beta)/(\alpha-1) > 0 \qquad (k = 0, 1, 2, \cdots).$$

Also,

$$\Delta^n M_k = (\alpha - \beta)/(\alpha - 1) \cdot \Delta^n (1/a_k) \geqq 0 \quad (n = 1, 2, 3, \cdots; k = 0, 1, 2, \cdots),$$

since $\{1/a_n\}$ is totally monotone for $\alpha > 1$. Hence $\{M_n\}$ is totally monotone and, therefore, Theorem 4 follows.

For $0 < \beta < \alpha < 1$,

$$1/M_n = (1 - \alpha)/(1 - \beta) + (\alpha - \beta)/(1 - \beta) \cdot 1/b_n.$$

Now $1/b_n > 0$; but $\{1/b_n\}$ is not totally monotone for $0 < \beta < 1$ by Theorem 2(i); consequently, $\{1/M_n\}$ is not totally monotone and the proof of Theorem 5 is complete.

COROLLARY. Taking $\beta = 0$ and $\mu(t) = 1/(t+1)$ we have from the first part of Theorem 4, with the notations of (2), M t.s. S_α, a result already well known ([11] Loc. Cit.; it will follow from the proof of Theorem 6 below that this result may be extended to the case of $0 < \alpha < 1$).

Proof of Theorem 6. For the proof of the first part of Theorem 6 it is sufficient to show that for given α, β with $0 \leqq \beta < \alpha < 1$, there is a totally monotone restricted moment function $\mu(t)$ for which $\bigcup_\beta$ t.s. $\bigcup_\alpha$. We take $\mu(t) = 1/(t+1)$ $(t \geqq 0)$, the moment function for M, which is obviously restricted. Then $M(t)$ being the moment function for $\bigcup_\beta / \bigcup_\alpha$,

$$M(t) = (1 - \beta)/(1 - \alpha) - (\alpha - \beta)/(1 - \alpha) \cdot (t+1)/(\alpha t + 1) > 0;$$

and

$$(-1)^n M^{(n)}(t) = n!(\alpha - \beta)\alpha^{n-1}/(\alpha t + 1)^{n+1} > 0 \qquad (n = 1, 2, 3, \cdots).$$

Thus $M(t)$ is totally monotone and accordingly,

$$\bigcup_\beta \text{ t.s. } \bigcup_\alpha.$$

For the proof of the second part of the Theorem we require one more lemma.

LEMMA 5. *If $M(t)$ is totally monotone for $t \geqq 0$, then, for $t > 0$,*

$$F(t) \equiv M'(t)M'''(t) - \{M''(t)\}^2 \geqq 0.$$

([17], p. 167)

To prove (ii) we take $H = \frac{1}{2}(E + E_e)$, a totally regular matrix, E_e being the Euler matrix of order e, so that $\mu(t) = \frac{1}{2}(1 + e^{-t})$ $(t \geqq 0)$. Then

$$\mu(x + iy) = \tfrac{1}{2}\{1 + e^{-(x+iy)}\} = \tfrac{1}{2}\{1 + e^{-x}(\cos y - i \sin y)\},$$

so that

$$\Re\mu(t) = \tfrac{1}{2}\{1 + e^{-x} \cos y\} \geqq 0 \text{ for } x \geqq 0.$$

$\mu(t)$ is, therefore, a restricted moment function. We have now

$$M(t) = (1-\beta)/(1-\alpha) - 2(\alpha-\beta)/(1-\alpha) \cdot e^t/\{1-\alpha+(1+\alpha)e^t\}.$$

As such

$$M'(t) = -2(\alpha-\beta)e^t/[(1-\alpha)+(1+\alpha)e^t]^2,$$

$$M''(t) = -2(\alpha-\beta) \cdot \{(1-\alpha)e^t - (1+\alpha)e^{2t}\}/[(1-\alpha)+(1+\alpha)e^t]^3.$$

$$M'''(t) = -2(\alpha-\beta) \cdot \{(1-\alpha)^2e^t - 4(1-\alpha^2)e^{2t} + (1+\alpha)^2e^{3t}\}/[(1-\alpha)+(1+\alpha)e^t]^4.$$

Hence

$$M'(0) = -\tfrac{1}{2}(\alpha-\beta),\ M''(0) = \alpha(\alpha-\beta)/2,\ M'''(0) = (\alpha-\beta)(1-3\alpha^2)/4.$$

Therefore $F(0) = (\alpha-\beta)^2(\alpha^2-1)/8 < 0$, as $0 < \alpha < 1$. Since $F(t)$ is continuous at $t = 0$, it follows that $F(t) < 0$ in a neighborhood $(0, \epsilon)$ on the right of the origin. Consequently, by Lemma 5, $M(t)$ cannot be totally monotone for $t \geqq 0$; and, therefore, $\bigcup_\beta$ n.t.s. $\bigcup_\alpha$.

PRESIDENCY COLLEGE, CALCUTTA.

REFERENCES.

[1] R. P. Agnew, "On equivalence of methods of evaluation of sequences," *Tôhoku Mathematical Journal*, vol. 35 (1932), pp. 244-252.

[2] ———, "On Hurwitz-Silverman-Hausdorff methods of summability," *Tôhoku Mathematical Journal*, vol. 49 (1942), pp. 1-14.

[3] S. K. Basu, "On the total relative strength of the Hölder and Cesàro methods," *Proceedings of the London Mathematical Society* (2), vol. 50 (1949), pp. 447-462.

[4] ———, "On the total relative strength of the Riesz and Cesàro methods," *Journal of the London Matematical Society*, vol. 24 (1949), pp. 51-59.

[5] W. H. J. Fuchs and W. W. Rogosinski, "On typical means," *Quarterly Journal of Mathematics*, vol. 14 (1943), pp. 27-48.

[6] G. H. Hardy, *Divergent Series* (Oxford, 1949).

[7] F. Hausdorff, "Summationsmethoden and Momentfolgen, I," *Mathematische Zeitschrift*, vol. 9 (1921), pp. 74-109.

[8] E. Hille and J. D. Tamarkin, "On moment functions," *Proceedings of the National Academy of Sciences*, vol. 19 (1933), pp. 902-908.

[9] W. A. Hurwitz, "Some properties of methods of evaluation of divergent sequences," *Proceedings of the London Mathematical Society* (2), vol. 26 (1927), pp. 231-248.

[10] E. R. Love, "Mercer's summability theorem," *Journal of the London Mathematical Society*, vol. 27 (1952), pp. 413-428.

[11] J. Mercer, "On the limits of real variants," *Proceedings of the London Mathematical Society* (2), vol. 5 (1907), pp. 206-224.

[12] H. R. Pitt, "Mercerian theorems," *Proceedings of the Cambridge Philosophical Society*, vol. 34 (1938), pp. 510-520.

[13] W. W. Rogosinski, "On Hausdorff's method of summability," *Proceedings of the Cambridge Philosophical Society*, vol. 38 (1942), pp. 166-192.

[14] ———, "On Hausdorff's methods of summability, II," *Proceedings of the Cambridge Philosophical Society*, vol. 38 (1942), pp. 344-363.

[15] E. C. Titchmarsh, *The Theory of Functions* (2nd Ed. Oxford, 1944).

[16] C. E. Walsh, "Note on an analogue of Mercer's theorem," *Journal of the London Mathematical Society*, vol. 17 (1942), pp. 13-17.

[17] D. V. Widder, *The Laplace Transform* (Princeton, 1946).

AN EMBEDDING THEOREM FOR SEMIGROUPS WITH CANCELLATION.* [1]

By TREVOR EVANS.

Introduction. The main part of this paper is concerned with the following question. Is there a finite integer n such that any countable cancellation semigroup can be embedded in a cancellation semigroup generated by n elements? [2] From other theorems of this type (see [1], [2], [4], [6]) one is tempted to conjecture that $n = 2$. However, this is not the case and we give an example of a countable cancellation semigroup which cannot be embedded in a two-generator cancellation semigroup. We show, furthermore, that there cannot be, for any n, any embedding theorem for arbitrary cancellation semigroups of the same type as those obtained for groups, semigroups, loops, etc. in the papers quoted above. The difficulties arise from the existence of subgroups in the cancellation semigroup to be embedded and we are able to prove the following theorem. Any countable cancellation semigroup not containing sub-semigroups which are groups (other than the trivial group, possibly, if the cancellation semigroup has a unit) can be embedded in a two-generator cancellation semigroup.

As an application of the embedding theorem described above we show the existence of a finitely related two-generator cancellation semigroup which has an unsolvable word problem.

1. Preliminary ideas. A semigroup consists of a set of elements closed with respect to an associative multiplication. A cancellation semigroup is a semigroup in which $xy = xz$ or $yx = zx$ implies $y = z$. We need later the following definitions and simple lemmas, the first of which we state without proof.

LEMMA 1(i). *Let u, v be two elements in a cancellation semigroup such that $uv = v$ (or $vu = v$). Then u is a two-sided unit for all elements in the semigroup.*

* Received May 8, 1953.

[1] This work was done under contract with the Office of Naval Research.

[2] This problem was brought to my attention by Mr. J. Addison who remarked on the possible use of such an embedding in the study of certain decision problems for cancellation semigroups.

(ii) *Let S be a cancellation semigroup with unit e and let G be the set of all elements g in S such that $gh = e$ for some h in S. Then G is the unique maximal subgroup contained in S and the set of elements $S - G$ is an ideal of S.*

(iii) *Let S and G be as in* (ii). *If the product $u_1u_2 \cdots u_k$ of a finite set $u_1, u_2, \cdots, u_k$ of elements of S lies in G, then each u_i is in G.*

We now turn to semigroups and cancellation semigroups given by generators and relations. Consider all finite strings, or words, in the set of generators $g_1, g_2, g_3, \cdots$.[3] A set of equations $\rho_i = \sigma_i$, $1, 2, 3, \cdots$ between certain pairs of words is given and called the fundamental or defining relations. By a relation $\alpha = \beta$ between a pair of words α, β in the generators, we mean the following.

(1.1) $\alpha = \alpha$ is a relation, where α is any word in the generators,

(1.2) if $\alpha = \beta$ is a defining relation, then $\alpha = \beta$ is a relation,

(1.3) if $\alpha = \beta$ is a relation, then $\beta = \alpha$ is a relation,

(1.4) if $\alpha = \beta$ and $\beta = \gamma$ are relations, then $\alpha = \gamma$ is a relation,

(1.5) if $\alpha = \beta$ is a relation, then $g_i\alpha = g_i\beta$ and $\alpha g_i = \beta g_i$ are relations,

(1.6) if $g_i\alpha = g_i\beta$ or $\alpha g_i = \beta g_i$ is a relation, then $\alpha = \beta$ is a relation.

Two words α, β are defined to be equivalent if, $\alpha = \beta$ is a relation. The resulting equivalence classes $\{\alpha\}$ of words form a semigroup under the multiplication $\{\alpha\}\{\beta\} = \{\alpha\beta\}$ and this semigroup satisfies the cancellation laws. It is called the cancellation semigroup generated by $g_1, g_2, g_3, \cdots$ with defining relations $\rho_i = \sigma_i$, $i = 1, 2, 3, \cdots$.

In the above, if (1.6) is omitted, we obtain the definition of the semigroup generated by $g_1, g_2, g_3, \cdots$ with defining relations $\rho_i = \sigma_i$, $i = 1, 2, 3, \cdots$.

We now wish to introduce the notion of an elementary transformation of a word. Let S be a semigroup generated by $g_1, g_2, g_3, \cdots$ with defining relations $\rho_i = \sigma_i$, $i = 1, 2, 3, \cdots$. Let α, β be two words in S such that $\alpha \equiv \lambda\rho_i\mu$, $\beta \equiv \lambda\sigma_i\mu$ [4] where λ, μ are void or words in S. Then we say that β can be obtained from α by an elementary transformation and we shall sometimes write this as $\alpha \to \beta$ by $\rho_i \to \sigma_i$. Similarly, we say that α can be obtained

[3] It is immaterial whether or not we include the empty word 1 although, of course, if it is included, then the semigroup necessarily has a unit.

[4] The relation ($\equiv$) is used to denote identity between words.

from β by an elementary transformation and write $\beta \to \alpha$ by $\sigma_i \to \rho_i$. By (1.5), $\alpha = \beta$ is a relation in S. We have, however, a stronger result than this.

LEMMA 2. *Two words α, β in a semigroup S, given by generators and relations, are equivalent if, and only if, there is a finite sequence of words $\alpha, \alpha_1, \alpha_2, \cdots, \alpha_n, \beta$ in S such that each word in the sequence is obtained from the word immediately preceding it by an elementary transformation.*

Proof. For the proof that, if α, β are equivalent, then such a sequence exists, we refer to [5], Lemma 5. Assume then that such a sequence exists. By the preceding paragraph, $\alpha = \alpha_1, \alpha_1 = \alpha_2, \cdots, \alpha_n = \beta$ are relations in S and hence, by (1.4), $\alpha = \beta$ is a relation in S. That is, α and β are equivalent in S.

2. Two counter-examples. We give here two examples which illustrate the difficulties in obtaining an embedding theorem for arbitrary cancellation semigroups. These examples show that the embedding theorem obtained in the next section is, in a sense, the best possible.

Let K be the cancellation semigroup generated by $g_1, g_2, \cdots, g_{10}$ with defining relations.

$$(2.1) \qquad g_1 g_5 = g_2 g_6, \qquad g_3 g_5 = g_4 g_6, \qquad g_1 g_7 = g_2 g_8, \qquad g_9 g_{10} = 1.$$

We note two properties of K. It cannot be embedded in a group since the subsemigroup H generated by $g_1, g_2, \cdots, g_8$ is the cancellation semigroup constructed by Malcev [3] which cannot be embedded in a group. Furthermore, g_9 generates an infinite cyclic group.

Let us assume that a cancellation semigroup C exists, generated by two elements a, b, such that C contains a subsemigroup K' isomorphic to K. Since K' has a unit, by Lemma 1 this is a unit for all elements of C. There will be no ambiguity if we denote this unit by 1 also. Let u, v be the elements of K' which correspond to g_9, g_{10} respectively in the isomorphism between K' and K. Since $g_9, g_{10} = 1$ in K, $uv = 1$ in C and so u and v are contained in a subgroup G of C. Now u, v as elements of C are generated by a and b. There are three cases to consider.

(i) One, say u, can be generated by a alone, and the other can be generated by b alone. Then $u = a^m$, $v = b^n$. Since u, v are in G, Lemma 1(iii) implies that a and b are in G. Hence C is a group.

(ii) One, say u, can be written as a product involving both a and b. Again, by Lemma 1(iii) this implies that a and b are in G and hence that C is a group.

(iii) Both u and v can be generated by one of the generators of F, say a. Then $u = a^m$, $v = a^n$. Since $uv = 1$, we have $a^{m+n} = 1$ in C. Hence $u^{m+n} = a^{(m+n)m} = 1$.

In cases (i) and (ii) we obtain a contradiction since K cannot be embedded in a group. In case (ii) we obtain a contradiction since $g_9{}^{m+n} \neq 1$ in K. Hence K cannot be embedded in a cancellation semigroup generated by two elements.

We may now ask whether there is, for some n greater than 2, an embedding theorem for arbitrary cancellation semigroups similar to the theorems obtained for other systems in [1], [2], [4], [6]. Assume that such a theorem holds. It will be as follows.

Let S be a cancellation semigroup generated by $g_1, g_2, g_3, \cdots$ with defining relations $\rho_i(g_1, g_2, g_3, \cdots) = \sigma_i(g_1, g_2, g_3, \cdots)$ $i = 1, 2, 3, \cdots$. Then there is an infinite set of words $\theta_1, \theta_2, \theta_3, \cdots$ in $a_1, a_2, a_3, \cdots, a_n$ such that in the cancellation semigroup F generated by $a_1, a_2, \cdots, a_n$ with defining relations $\rho_i(\theta_1, \theta_2, \theta_3, \cdots) = \sigma_i(\theta_1, \theta_2, \theta_3 \cdots)$, $i = 1, 2, 3, \cdots$ the subsemigroup generated by $\theta_1, \theta_2, \theta_3, \cdots$ is isomorphic to S.

We note in the above that since $\theta_1, \theta_2, \theta_3, \cdots$ is an infinite set of words in a finite set of symbols $a_1, a_2, a_3, \cdots, a_n$, there is an infinite subset $\theta_{i(1)}$, $\theta_{i(2)}, \theta_{i(3)}, \cdots$ of the θ_i and a finite subset $a_{j(1)}, a_{j(2)}, \cdots, a_{j(k)}$ of the a_j such that each θ_i in $\theta_{i(1)}, \theta_{i(2)}, \theta_{i(3)}, \cdots$ is a word containing all symbols in the set $a_{j(1)}, a_{j(2)}, \cdots, a_{j(k)}$ and no other a_j.

We now attempt to apply the above embedding theorem to a cancellation semigroup consisting of the generators and relations of K in the previous example and an infinite number of free generators. If in this embedding the generators $g_1, g_2, \cdots, g_{10}$ of K correspond to θ_i in the set $\theta_{i(1)}, \theta_{i(2)}, \theta_{i(3)}, \cdots$, then $a_{j(1)}, a_{j(2)}, \cdots, a_{j(k)}$ will all have inverses in F and so K will be embedded in a group.

From this contradiction, it follows that the embedding theorem stated above does not hold and that if we are to formulate an embedding theorem for arbitrary cancellation semigroup S, we must be able to distinguish between those generators of S which have inverses and those which do not. However, in an arbitrary cancellation semigroup given by generators and relations, there is probably no general method of deciding whether an element in the semigroup has an inverse. Hence there will be no general method of embedding an arbitrary cancellation semigroup in an n-generator cancellation semigroup.

This does not mean that we cannot show that any countable cancellation semigroup can be embedded in a cancellation semigroup generated by n elements for some n. In fact, we conjecture that $n = 4$. The reasons for

this are as follows. A cancellation semigroup, by Lemma 1, can be split into two disjoint parts, one a group G and the other, a cancellation semigroup H not containing elements with inverses. We can embed G in a group generated by two elements of finite order [5] that is, in a cancellation semigroup A generated by two elements, and using the theorem proved in the next section, we can embed H in a cancellation semigroup B generated by two elements. Further relations must now be added to the elements of A and B corresponding to the relations between the elements of G and H in S. If these relations do not affect the embedding of G in A and H in B, then S is embedded in a cancellation semigroup generated by four elements. This method certainly avoids the difficulties of the above examples and it would be interesting to know whether it does, in fact, give an embedding of S, especially since the explicit construction of the embedding for a given cancellation semigroup requires a method for determining which generators of S have inverses and, as we have remarked above, such a method probably does not exist.

3. The embedding theorem. Let S be a cancellation semigroup generated by a finite or countably infinite set of generators $g_1, g_2, g_3, \cdots$ and defined by a finite or countably infinite set of relations

$$\rho_i(g_1, g_2, g_3, \cdots) = \sigma_i(g_1, g_2, g_3, \cdots), \qquad i = 1, 2, 3, \cdots. \tag{3.1}$$

We shall assume that S does not contain any subgroups. By Lemma 1, this is equivalent to assuming that there is no relation in S of the form $g_i\alpha = g_i$ or $\alpha g_i = g_i$ where α is a non-empty word in the generators.[6]

We construct now a semigroup F generated by two elements a and b. Let θ_i denote the word $b^i a b^i a^2 b^i$, $i = 1, 2, 3, \cdots$; and $\mathcal{K}$ the set of words generated by $\theta_1, \theta_2, \theta_3, \cdots$. There is a one-one correspondence between the words in S and the words in $\mathcal{K}$ obtained by making g_i correspond to θ_i. If α is any word in S, α' will denote the corresponding word in $\mathcal{K}$. We also use the following notational device. If the word α in F is of the form $b^i\gamma$, then $b^{-j}\alpha$ will denote the word $b^{i-j}\gamma$ for $j < i$ and $b^{-i}\alpha$ will denote the word γ. A similar notation will be used for words ending with a power of b.

The defining relations of F can now be described. Consider all relations $\alpha = \beta$ in S such that α and β do not begin with the same generator nor end

[5] This result (unpublished), an extension of one of the results in [2], has been obtained by B. H. Neumann.

[6] Here, and throughout this section, we do not include the empty word as a word in S.

with the same generator. We call such relations reduced. These relations have one of the following forms:

$$(3.2) \qquad g_i\gamma g_j = g_k\delta g_m,\ i \neq k,\ j \neq m;\ g_i\gamma g_j = g_k,\ i, j \neq k;\ g_i = g_j,\ i \neq j.$$

Here γ and δ are words in S or possibly void.

Corresponding to every reduced relation $\alpha = \beta$ in S of the form $g_i\gamma g_j = g_k\delta g_m$ we take $b^{-p}\alpha' b^{-q} = b^{-p}\beta' b^{-q}$ as a defining relation of F where $p = \min(i, k)$, $q = \min(j, m)$.

Corresponding to every reduced relation $\alpha = \beta$ in S of the form $g_i\gamma g_j = g_k$ we take $b^{-p}\alpha' b^{-q} = b^{-p}\beta' b^{-q}$ as a defining relation of F where $p = \min(i, k)$, $q = \min(j, k)$. Here, of course, β' is θ_k.

Corresponding to every reduced relation $\alpha = \beta$ in S of the form $g_i = g_j$ we take $b^{-p}\alpha' b^{-p} = b^{-p}\beta' b^{-p}$ as a defining relation of F where $p = \min(i, j)$. Here, α' and β' are θ_i and θ_j respectively.

As an example, suppose $g_1g_2g_3 = g_4g_2$ is a relation in S. Then $aba^2b\theta_2b^3ab^3a^2b = b^3ab^4a^2b^4b^2ab^2a^2$ would be a defining relation of F.

In what follows we will constantly be using special properties of words in $\mathcal{K}$. These properties are specified in the next lemma.

LEMMA 3. *Let ω be a word in $\mathcal{K}$ of the form $\lambda\pi\mu$ (λ, μ may be void). If the subword π begins with $b^{r-p}ab^ra^2$ (for some r, $0 \leq p \leq r$), and ends with $ab^sa^2b^{s-q}$ (for some s, $0 \leq q \leq s$), then $b^p\pi b^q$ is a word in $\mathcal{K}$ and λb^{-p}, $b^{-q}\mu$ are void or words in $\mathcal{K}$.*

Proof. A word in $\mathcal{K}$ has the form

$$b^iab^ia^2b^{i+j}ab^ja^2b^{j+\cdots}\cdots b^{\cdot+m}ab^ma^2b^{m+n}ab^na^2b^n.$$

The lemma follows from this. We leave the details of proof to the reader.

LEMMA 4(i). *If $\alpha = \beta$ is a relation in S, then $\alpha' = \beta'$ is a relation in F,* (ii) *if $\zeta = \eta$ is a relation in F and if ζ is a word in $\mathcal{K}$, then η is also a word in $\mathcal{K}$,* (iii) *if α', β' are two words in $\mathcal{K}$ such that $\alpha' = \beta'$ is a relation in F, then $\alpha = \beta$ is a relation in S.*

Proof. (i) Let $\alpha = \beta$ be a relation in S. This relation can be written as $\lambda\phi\mu = \lambda\psi\mu$ where ϕ, ψ do not begin with the same generator nor end with the same generator and λ, μ are void or words in S. Then $\phi = \psi$ is a reduced relation in S and, for some p, q, $b^{-p}\phi' b^{-q} = b^{-p}\psi' b^{-q}$ is a defining relation of F. Since $\alpha' \equiv \lambda' b^p b^{-p}\phi' b^{-q} b^q\mu'$, $\beta' \equiv \lambda' b^p b^{-p}\psi' b^{-q} b^q\mu'$, it follows that $\alpha' = \beta'$ is a relation in F.

(ii) One sees from Lemma 3 that if a word ζ in $\mathcal{K}$ admits an elementary transformation, then $\zeta \equiv \lambda' b^p b^{-p} \phi' b^{-q} b^q \mu'$ where λ', μ' are words in $\mathcal{K}$ or void and $b^{-p}\phi' b^{-q} = b^{-p}\psi' b^{-q}$ is a defining relation of F.

If $\zeta = \eta$ is a relation in F, by Lemma 3 there is a finite sequence of elementary transformations carrying ζ into η. Assume, first of all, that this sequence consists of just one transformation. By the preceding paragraph $\zeta \equiv \lambda' b^p b^{-p} \phi' b^{-q} b^q \mu'$ and so $\eta \equiv \lambda' b^p b^{-p} \psi' b^{-q} b^q \mu'$. Hence $\eta \equiv \lambda' \psi' \mu'$ is in $\mathcal{K}$. A simple induction on the number of elementary transformations connecting ζ and η completes the proof.

(iii) If α', β' are connected by one elementary transformation, then, by part (ii) of this lemma, $\alpha' \equiv \lambda' \phi' \mu'$, $\beta' \equiv \lambda' \psi' \mu'$ where λ', μ' are in $\mathcal{K}$ and $\phi = \psi$ is a relation in S. Then $\alpha \equiv \lambda \phi \mu$, $\beta \equiv \lambda \psi \mu$ and so $\alpha = \beta$ is a relation in S. In general, we have a finite sequence of words $\alpha', \alpha'_1, \alpha'_2, \cdots, \alpha'_n, \beta'$ all in $\mathcal{K}$, such that each word is obtained from the word preceding it by an elementary transformation. Then $\alpha = \alpha_1, \alpha_1 = \alpha_2, \cdots, \alpha_n = \beta$ are relations in S and so $\alpha = \beta$ is a relation in S.

LEMMA 5. *The subsemigroup of F generated by $\theta_1, \theta_2, \theta_3, \cdots$ is isomorphic to S.*

Proof. By Lemma 4 two equivalent words in F are either both in $\mathcal{K}$ or both not in $\mathcal{K}$. The equivalence classes consisting of word in $\mathcal{K}$ form a subsemigroup K of F. There is a one-one correspondence $\alpha \leftrightarrow \alpha'$ between words α in S and words α' in K. Now, by Lemma 4, two words α, β in S are equivalent if, and only if, α' and β' are equivalent in F. Hence, the semigroups S and K are isomorphic.

We have now embedded S in a semigroup generated by two elements. The next step is to show that F can be regarded as a cancellation semigroup and to do this we need to obtain some properties of the defining relations of F.

Let ω be a word in F and let $\omega \to \omega_1$ by $\zeta_1 \to \eta_1$, $\omega_1 \to \omega_2$ by $\zeta_2 \to \eta_2$ where $\zeta_1 \to \eta_1$, $\zeta_2 \to \eta_2$ are elementary transformations of F. We say that these two successive elementary transformations of ω overlap if part of ω_1 lies both in the η_1 of the first transformation and in the ζ_2 of the second transformation. There are four ways in which this can happen.

(a) ζ_2 has the form $\lambda \eta_1 \mu$. Then

$$\omega \equiv \kappa \lambda \zeta_1 \mu \tau \to \omega_1 \equiv \kappa \lambda \eta_1 \mu \tau \to \omega_2 \equiv \kappa \eta_2 \tau \tag{3.3}$$

(b) ζ_2 has the form $\lambda \pi$, η_1 has the form $\pi \mu$. Then

(3. 4) $$\omega \equiv \kappa\lambda\zeta_1\tau \rightarrow \omega_1 \equiv \kappa\lambda\pi\mu\tau \rightarrow \omega_2 \equiv \kappa\eta_2\mu\tau$$

(c) η_1 has the form $\lambda\zeta_2\mu$. Then

(3. 5) $$\omega \equiv \kappa\zeta_1\tau \rightarrow \omega_1 \equiv \kappa\lambda\zeta_2\mu\tau \rightarrow \omega_2 \equiv \kappa\lambda\eta_2\mu\tau$$

(d) η_1 has the form $\lambda\pi$, ζ_2 has the form $\pi\mu$. Then

(3. 6) $$\omega \equiv \kappa\zeta_1\mu\tau \rightarrow \omega_1 \equiv \kappa\lambda\pi\mu\tau \rightarrow \omega_2 \equiv \kappa\lambda\eta_2\tau.$$

In all these cases λ, μ, κ, τ are void or words in F. The word π is assumed non-empty since, otherwise, the two transformations would not overlap.

The overlap, that is, the subword of ω_1 which is in both η_1 and ζ_2, is $\eta_1, \pi, \zeta_2, \pi$ in cases (a), (b), (c), (d) respectively. If this overlap γ is such that, for some p, q, $b^p\gamma b^q$ is a word in $\mathcal{K}$, then we say that the transformations $\mathcal{K}$-overlap. Since, for some m, n, p, q, $b^m\eta_1 b^n$, $b^p\zeta_2 b^q$ are in $\mathcal{K}$, the overlaps in (a) and (c) are always $\mathcal{K}$-overlaps. In cases (b) and (d) the situation is a little more complicated. We will restrict our discussion to (d) since (b) can be treated in the same way.

Since $\lambda\pi$ is one side of a defining relation it is of the form $\phi ab^n a^2$ or $\phi ab^s a^2 b^{s-n}$, where ϕ is such that $b^m\lambda\pi b^n$ is in $\mathcal{K}$. Similarly, $\pi\mu$ can be written as $ab^p a^2\psi$ or $b^{r-p}ab^r a^2\psi$. In the following cases we shall say that the transformations have a short overlap of type (d).

(3. 7) $\pi \equiv a$, $\lambda \equiv \phi ab^n a$, $\mu \equiv b^p a^2\psi$,

(3. 8) $\pi \equiv ab^p$, $\lambda \equiv \phi ab^s a$, $\mu \equiv a^2\psi$, $(p = s - n$ in this case),

(3. 9) $\pi \equiv b^i$, $\lambda \equiv \phi ab^s a^2 b^{s-n-i}$, $\mu \equiv b^{r-p-i}ab^r a^2\psi$, $(i \leq \min(s - n, r - p))$.

An examination of the beginning of $\pi\mu$ and the end of $\lambda\pi$ shows that it is not possible for π to be $b^{r-p}a$, $b^{r-p}ab^i$, $b^{r-p}ab^r a$, $ab^p a$, $a^2 b^{s-n}$, $b^i a^2$ or a^2. It follows that, except for the short overlappings listed above, π necessarily must begin with $ab^p a^2$ or $b^{r-p}ab^r a^2$ and end with $ab^n a^2$ or $ab^s a^2 b^{s-n}$. In this case, since $b^m\lambda\pi b^n$, $b^p\pi\mu b^q \in \mathcal{K}$, by Lemma 3, $b^p\pi b^n \in \mathcal{K}$. Hence, except for the short overlaps, all overlaps of type (d) are $\mathcal{K}$-overlaps.

Similarly, except for three short overlaps, all overlaps of type (b) are $\mathcal{K}$-overlaps.

LEMMA 6. *Let $\omega \rightarrow \omega_1 \rightarrow \omega_2$ by $\zeta_1 \rightarrow \eta$, $\zeta_2 \rightarrow \eta_2$ respectively, the two transformation having a $\mathcal{K}$-overlap. Then there is an elementary transformation $\zeta_3 \rightarrow \eta_3$ of F such that $\omega \rightarrow \omega_2$ by $\zeta_3 \rightarrow \eta_3$.*

Proof. The method of proof will be sufficiently illustrated if we consider overlaps of type (d) only.

For some m, n, p, q, we have

$$(3.10) \qquad b^m\zeta_1 b^n \equiv \alpha'_1,\ b^m\lambda\pi b^n \equiv \beta'_1,\ b^p\pi\mu b^q \equiv \alpha'_2,\ b^p\eta_2 b^q \equiv \beta'_2,$$

where $\alpha_1 = \beta_1$, $\alpha_2 = \beta_2$ are relations in S. Thus both the words $b^m\lambda\pi b^n$, $b^p\pi\mu b^q$ are in $\mathcal{K}$. As we have noted above $b^p\pi b^n$ is in $\mathcal{K}$ since the overlap is a $\mathcal{K}$-overlap. Considering $b^p\pi b^n$ as a subword of $b^m\lambda\pi b^n$, by Lemma 3, the word $b^m\lambda b^{-p}$ is either a word in $\mathcal{K}$ or possibly void. Similarly $b^{-n}\mu b^q$ is either a word in $\mathcal{K}$ or void.

Hence, in the relations $\alpha' b^{-n}\mu b^q = \beta'_1 b^{-n}\mu b^q$, $b^m\lambda b^{-p}\alpha'_2 = b^m\lambda b^{-p}\beta'_2$, both sides are words in $\mathcal{K}$. These two relations can be written as

$$(3.11) \qquad b^m\zeta_1\mu b^q = b^m\lambda\pi\mu b^q, \quad b^m\lambda\pi\mu b^q = b^m\lambda\eta_2 b^q.$$

We obtain the new relation $b^m\zeta_1\mu b^q = b^m\lambda\eta_2 b^q$ and since both sides of this relation are in $\mathcal{K}$, by Lemma 4, there is a relation $\alpha_3 = \beta_3$ in S such that $\alpha'_3 \equiv b^m\zeta_1\mu b^q$, $\beta'_3 \equiv b^m\lambda\eta_2 b^q$. Now α_1, α_3 begin with the same generator, and β_1, β_3 begin with the same generator. Hence α_3, β_3 do not begin with the same generator. Similarly, α_3, β_3 do not end with the same generator. It follows that $b^{-m}\alpha'_3 b^{-q} = b^{-m}\beta'_3 b^{-q}$ or $\zeta_1\mu = \lambda\eta_2$ is a defining relation of F, giving the required $\zeta_3 \to \eta_3$ since $\omega \equiv \kappa\zeta_1\mu\tau$, $\omega_2 \equiv \kappa\lambda\eta_2\tau$.

The remaining cases are proved in the same way, i. e. by considering those relations in S which correspond to the overlapping transformations in F.

LEMMA 7. *Let $a\alpha$ and $a\beta$ be two equivalent words in F. Then α and β are equivalent in F.*

Proof. The proof is by induction on the number of elementary transformations connecting $a\alpha$ and $a\beta$. Consider first of all the case where $a\alpha \to a\beta$ by $\zeta \to \eta$. As a defining relation of F one side of $\zeta = \eta$ begins with a and the other side with a power of b. Hence $\zeta \to \eta$ cannot affect the beginning of $a\alpha$ since both $a\alpha$ and $a\beta$ begin with a. The transformation must, therefore, occur inside α, carrying α into β. That is, α and β are equivalent.

Assume now that the lemma is true whenever $a\alpha$ can be transformed into $a\beta$ by n or fewer elementary transformations and consider the case where $a\alpha$ and $a\beta$ are connected by $n+1$ transformations. We will denote this by

$$(3.12) \qquad a\alpha \to \omega_1 \to \omega_2 \to \cdots \to \omega_n \to a\beta$$

the succesive transformations being $\zeta_1 \to \eta_1$, $\zeta_2 \to \eta_2$, $\cdots$, $\zeta_{n+1} \to \eta_{n+1}$. We can dispose of three cases immediately.

(i) Suppose that, for some i, ω_i begins with a, that is ω_i is of the form

$a\gamma$. Then, by the inductive hypothesis, both α and β are equivalent to γ. Hence α and β are equivalent.

(ii) Suppose that, for some i, the transformations $\zeta_i \to \eta_i, \zeta_{i+1} \to \eta_{i+1}$ have a $\mathcal{K}$-overlap. By Lemma 6 these two transformations can be replaced by one. Hence $a\alpha$ and $a\beta$ can be connected by n elementary transformations and so α and β are equivalent.

(iii) Suppose that $a\alpha$ can be written as $\kappa\tau$ and that the first i transformations all occur in κ; that is, τ is not affected by any of the transformations $\zeta_1 \to \eta_1, \cdots, \zeta_i \to \eta_i$. Let the transformation $\zeta_{i+1} \to \eta_{i+1}$ occur in τ. Using an obvious notation, we can write the first $i+1$ transformations as

$$a\alpha \to \kappa_1\tau \to \kappa_2\tau \to \cdots \to \kappa_i\tau \to \kappa_i\tau_1. \tag{3.13}$$

But then the sequence of transformations

$$a\alpha \to \kappa\tau_1 \to \kappa_1\tau_1 \to \kappa_2\tau_1 \to \cdots \to \kappa_i\tau_i \tag{3.14}$$

will also transform $a\alpha$ into ω_{i+1}, and so the sequence of $n+1$ transformations

$$a\alpha \to \kappa\tau_1 \to \cdots \to \kappa_i\tau_1 \to \omega_{i+2} \to \cdots \to \omega_n \to a\beta \tag{3.15}$$

transforms $a\alpha$ into $a\beta$. Now in this last sequence, the second term $\kappa\tau_1$ begins with a since $a\alpha \equiv \kappa\tau$. Hence, by (i), α and β are equivalent.

We shall assume from this point that in the sequence (3.12) the three situations described above do not occur. Then $\zeta_1 \to \eta_1$ must occur at the left end of $a\alpha$ and $\zeta_1 \to \eta$, $\zeta_2 \to \eta_2$ have a short overlap of type (d). The next transformation $\zeta_3 \to \eta_3$ must then either have a short overlap of type (b) or (d) with $\zeta_2 \to \eta_2$ or ζ_3 must lie to the left of η_2 in ω_2.

Similarly, if $\zeta_2 \to \eta_2$, $\zeta_3 \to \eta_3$ do have a short overlap of type (d), then either $\zeta_3 \to \eta_3$, $\zeta_4 \to \eta_4$ have a short overlap of type (b) or (d), or ζ_4 must lie to the left of η_3 in ω_3. In general, if each pair of successive transformations $\zeta_i \to \eta_i, \zeta_{i+1} \to \eta_{i+1}$, $i = 1, 2, \cdots, k$, have a short overlap of type (d), then $\zeta_{k+1} \to \eta_{k+1}$, $\zeta_{k+2} \to \eta_{k+2}$ either have a short overlap of type (b) or (d), or ζ_{k+2} must lie to the left of η_{k+1} in ω_{k+1}.

Let r be the number such that each pair of successive transformations in $\zeta_1 \to \eta_1, \zeta_2 \to \eta_2, \cdots, \zeta_r \to \eta_r$, has a short overlap of type (d) but $\zeta_r \to \eta_r$, $\zeta_{r+1} \to \eta_{r+1}$ do not have a short overlap of type (d). We have seen above that $r \geq 2$. Since ω_n begins with b, the elementary transformation taking ω_n into $a\beta$, cannot have a short overlap of type (d) with the preceding transformation. Hence, $r \leq n$.

We can therefore write the first r transformations as

$$(3.16)\quad \zeta_1 \to \lambda_1\pi_1,\ \pi_1\mu_1 \to \lambda_2\pi_2,\ \pi_2\mu_2 \to \lambda_3\pi_3, \cdots,$$
$$\pi_{r-2}\mu_{r-2} \to \lambda_{r-1}\pi_{r-1},\ \pi_{r-1}\mu_{r-1} \to \eta_r$$

where the short overlaps $\pi_1, \pi_2, \cdots, \pi_{r-1}$, are as listed in (3.7), (3.8), (3.9). Then

$$(3.17)\quad a\alpha \equiv \zeta_1\mu_1\mu_2 \cdots \mu_{r-1}\chi,$$
$$\omega_1 \equiv \lambda_1\pi_1\mu_1\mu_2 \cdots \mu_{r-1}\chi, \cdots, \omega_r \equiv \lambda_1\lambda_2 \cdots \lambda_{r-1}\eta_r\chi$$

and the transformation $\zeta_{r+1} \to \eta_{r+1}$ occurs in $\lambda_1\lambda_2 \cdots \lambda_{r-1}\eta_r$.

We consider now in what way ζ_{r+1} can be a subword of $\lambda_1\lambda_2 \cdots \lambda_{r-1}\eta_r$ with the restriction that it cannot have a $\mathcal{K}$-overlap with η_r. An examination of the form of λ_i in (3.7), (3.8), (3.9) coupled with the fact that, for some j, k, $b^j\zeta_{r+1}b^k$ is a word in $\mathcal{K}$, shows that only two cases are possible.

(1) ζ_{r+1} is a subword of some λ_t, $1 \leq t \leq r-1$. Let λ_t go into $\bar{\lambda}_t$ under the transformation $\zeta_{r+1} \to \eta_{r+1}$. Then it is easily verified that the sequence of r elementary transformations (3.16) with λ_t replaced by $\bar{\lambda}_t$ transforms $a\alpha$ into ω_{r+1}. Hence $a\alpha$ can be transformed into $a\beta$ by n elementary transformations and the lemma is proved.

(2) Here we consider those cases in which ζ_{r+1} does not lie entirely in one of the λ_i. If ζ_{r+1} contains subwords of more than one λ_i, it must contain the right end of some λ_t and the left end of λ_{t+1} (or η_r), for $1 \leq t \leq r-1$. We use now the fact that if one side of a defining relation of F begins with a, then the other side begins with a power of b. In (3.7), if λ_t ends with ab^na, then π_t is a and hence λ_{t+1} (or η_r) begins with a power of b. In this case ζ_{r+1} cannot overlap both λ_t and λ_{t+1} (or η_r). In (3.8), if λ_t ends in ab^sa, then π_t is ab^p and hence λ_{t+1} (or η_r) begins with a power of b. In this case ζ_{r+1} cannot overlap both λ_t and λ_{t+1} (or η_r). In (3.9), if λ_t ends in $ab^sa^2b^{s-n-i}$, then π_t is b^i and hence λ_{t+1} (or η_r) begins with a. In this case ζ_{r+1} can overlap λ_t by a power of b, the rest of ζ_{r+1} being contained in λ_{t+1} (we can disregard η_r here since we have assumed that $\zeta_{r+1} \to \eta_{r+1}$ does not have a $\mathcal{K}$-overlap with $\zeta_r \to \eta_r$).

In this final case, all of ζ_{r+1} except a power of b at the left end lies in λ_{t+1}. Hence the transformations $\pi_t\mu_t \to \lambda_{t+1}\pi_{t+1}$ and $\zeta_{r+1} \to \eta_{r+1}$ have a $\mathcal{K}$-overlap. Now

$$(3.18)\quad \omega_{t+1} \equiv \lambda_1\lambda_2 \cdots \lambda_t\lambda_{t+1}\pi_{t+1}\mu_{t+1} \cdots \mu_{r-1}\chi$$

where the transformations $\zeta_{t+2} \to \eta_{t+2}, \cdots, \zeta_r \to \eta_r$ all occur in the subword

$\pi_{t+1}\mu_{t+1}\cdots\mu_{r-1}$. Since $\zeta_{r+1}\rightarrow\eta_{r+1}$ occurs in the subword $\lambda_t\lambda_{t+1}$, it is easily seen that the following sequences of $r+1$ elementary transformations

$$(3.19)\quad \zeta_1\rightarrow\eta_1,\cdots,\zeta_{t+1}\rightarrow\eta_{t+1},\zeta_{r+1}\rightarrow\eta_{r+1},\zeta_{t+2}\rightarrow\eta_{t+2},\cdots,\zeta_r\rightarrow\eta_r$$

transforms $a\alpha$ into ω_{r+1}. Replacing $\zeta_{t+1}\rightarrow\eta_{t+1},\zeta_{r+1}\rightarrow\eta_{r+1}$ by a single elementary transformation, as we may do since these two transformations have a $\mathcal{K}$-overlap, we can transform $a\alpha$ into ω_{r+1} by r elementary transformations and hence $a\alpha$ into $a\beta$ by n elementary transformations. Then, by our inductive hypothesis, α and β are equivalent. This completes the proof of the lemma.

Lemma 7 requires only trivial changes in order to prove that if $b\alpha$ and $b\beta$ are equivalent words in F, then α and β are equivalent. Also, if in the above, we consider short overlaps of type (b), the same method of proof shows that if αa and βa (αb and βb) are equivalent words in F, then α and β are equivalent.

Hence if $g\alpha=g\beta$ or $\alpha g=\beta g$ (g is a generator, a or b) is a relation in F, then $\alpha=\beta$ is a relation in F. It follows that two words are equivalent in F if and only if they are equivalent in the cancellation semigroup C having the same generators and defining relations as F. Thus, the identity mapping is an isomorphism between F and C. We have now embedded the cancellation semigroup S in a cancellation semigroup C generated by two elements. However, the defining relations of C are not in a very useful form. This is remedied by the next lemma.

LEMMA 8. *The cancellation semigroup C (and hence the semigroup F) is isomorphic to the cancellation semigroup D generated by a, b with defining relations*

$$(3.20)\quad \rho_i(\theta_1,\theta_2,\theta_3,\cdots)=\sigma_i(\theta_1,\theta_2,\theta_3,\cdots),\qquad i=1,2,\cdots.$$

Proof. It is sufficient to show that C satisfies the defining relations of D and that D satisfies the defining relations of C. Now $\rho_i(g_1,g_2,g_3,\cdots)=\sigma_i(g_1,g_2,g_3,\cdots)$ is a relation in S and so, by Lemma 4(i), $\rho_i(\theta_1,\theta_2,\theta_3,\cdots)=\sigma_i(\theta_1,\theta_2,\theta_3,\cdots)$ is a relation in F, therefore a relation in C. Hence C satisfies the defining relations of D. Furthermore, a defining relation of C has the form $b^{-p}\alpha' b^{-q}=b^{-p}\beta' b^{-q}$ where $\alpha=\beta$ is a relation in S. It is easy to show that if $\alpha=\beta$ is a relation in S then $\alpha'=\beta'$ is a relation in D.[7] On cancelling powers of b it follows that $b^{-p}\alpha' b^{-q}=b^{-p}\beta' b^{-q}$ is a relation in D. Hence D satisfies the defining relations of C.

We are now in a position to state the main theorem.

[7] Using the correspondence $g_i\leftrightarrow\theta_i$ we can copy in D the steps leading to the relation $\alpha=\beta$ in S.

THEOREM I. *Let S be a cancellation semigroup given by generators $g_1, g_2, g_3, \cdots$ and relations*

$$\rho_i(g_1, g_2, g_3, \cdots) = \sigma_i(g_1, g_2, g_3, \cdots), \qquad i = 1, 2, 3, \cdots$$

and such that S does not contain a group as a subsemigroup. Then S is isomorphic to the subsemigroup generated by $b^iab^ia^2b^i$, $i = 1, 2, 3, \cdots$, of the cancellation semigroup generated by a, b with defining relations

$$\rho_i(baba^2b, b^2ab^2a^2b^2, b^3ab^3a^2b^3, \cdots) = \sigma_i(baba^2b, b^2ab^2a^2b^2, b^3ab^3a^2b^3, \cdots), \qquad i = 1, 2, \cdots.$$

In the above work the empty word was excluded from consideration as a word in a semigroup or cancellation semigroup given by generators and relations. However, the proof of Theorem I goes through unchanged even if the empty words is included among the words of the semigroups considered. This will then be a unit element for S and the restriction, that for no non-empty α is $g_i\alpha = g_i$ or $\alpha g_i = g_i$ a relation, can be replaced by the following. There is no relation of the form $\alpha = 1$ in S, where α is a non-empty word in the generators of S.

COROLLARY. *A countable cancellation semigroup with a unit but without non-trivial subgroups can be embedded in a cancellation semigroup with a unit generated by two elements.*

As in [1], the above embedding theorem can be interpreted in terms of a property of certain subsemigroups of the free cancellation semigroup on two generators. In our construction we have used the subsemigroup generated by $\theta_1, \theta_2, \theta_3, \cdots$. We may ask which other semigroups can play the role of $\mathcal{K}$ in the proof of the embedding theorem. It is easy to obtain sets of generators to replace the θ_i which have a more complicated form than the θ_i as words in a, b; one such is $b^ia^ib^ia^ib^i$, $i = 1, 2, 3, \cdots$ but it seems likely that no simpler set of words can replace $\theta_1, \theta_2, \theta_3, \cdots$.

4. The word problem for cancellation semigroups.[8] A. M. Turing [5] has given an example of a finitely generated and related cancellation semigroup with an unsolvable word problem. Using the embedding of Theorem I, we can prove the following theorem.

---

[8] I am indebted to Dr. W. W. Boone for several illuminating discussions concerning the word problem for cancellation semigroups. The proof given that there is no relation $\alpha = 1$ (α a non-empty word) in Turing's semigroup is essentially due to Dr. Boone.

THEOREM II. *There exists a finitely related cancellation semigroup generated by two elements for which the word problem is unsolvable.*

Proof. It is only necessary to show that the cancellation semigroup constructed by Turing in [5] satisfies the condition of Theorem I necessary for it to be embedded in a two-generator cancellation semigroup. We shall use the terminology of [5] but adapt the notation to conform with the rest of this paper.

Let S be the semigroup given by the generators and defining relations on pp. 495-496 in [5] and let C be the cancellation semigroup with the same generators and defining relations. We first show that two words in the generators are equivalent in S if, and only if, they are equivalent in C. Turing has stated the equivalent of this in [5], p. 497, line 12 ff. but only proves that two normal words are equivalent in S if, and only if, they are equivalent in C ([5], Lemmas 17 and 19).

Any word α in the generators can be written as $\lambda g_1 g_2 \mu$ where $g_1 g_2$ is a barrier and λg_1 contains no barrier. Here λ and μ may be empty. Using Lemmas 5 to 11 in [5], in particular Lemma 11, if β is equivalent to α in S (i. e. β can be obtained by a sequence of elementary transformations from α), then β can be written in the form $\kappa g_3 g_4 \tau$ where $g_3 g_4$ is a barrier, λg_1 is equivalent to κg_3 in S, and $g_2 \mu$ is equivalent to $g_4 \tau$ in S. Furthermore κg_3 contains no barrier. Hence the transformation of α into β can be effected by first transforming λg_1 into κg_3, then transforming $g_2 \mu$ into $g_4 \tau$..

Assume now that α is of the form $g\gamma$ and β is of the form $g\delta$ where g is a generator. Then $\lambda g_1, \kappa g_3$ can be written as $g\rho, g\sigma$ respectively. As words without barriers, $g\rho$ and $g\sigma$ are either both normal or both consist entirely of left (right) symbols. If $g\rho, g\sigma$ are normal, by [5], Lemma 17, ρ and σ are equivalent in S. If $g\rho, g\sigma$ consist entirely of left (right) symbols, by [5], Lemma 18, the words $g\rho$ and $g\sigma$ are identical and so, again, ρ and σ are equivalent in S.

It follows that there is a sequence of elementary transformations taking γ into δ; i. e. γ and δ are equivalent in S. In the same way we can show that if two words γg, δg are equivalent in S, then γ and δ are equivalent in S. Thus two words are equivalent in S if, and only if, they are equivalent in C.

It is now easy to show that no relation of the form $\alpha = 1$ (or $g\alpha = g$ if we are not allowing the empty word) holds in C, where α is a non-empty word in the generators. For $\alpha = 1$ (or $g\alpha = g$) in C if, and only if, there is a sequence of elementary transformations of S carrying α into the empty word (or $g\alpha$ into g). Now, in the defining relations of S, no side of a relation

has length less than two in the generators. Hence, neither the empty word nor a word consisting of a single generator can be affected by an elementary transformation of S and so neither can be equivalent to any word other than itself.

The cancellation semigroup C therefore satisfies the condition necessary for Theorem I to apply and we can embed C in a cancellation semigroup generated by two elements. This two-generator cancellation semigroup has the same (finite) number of defining relations as C and obviously its word problem is unsolvable.

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] Trevor Evans, "Embedding theorems for multiplicative systems and projective geometries," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 614-620.

[2] Graham Higman, B. H. Neuman and Hanna Neumann, "Embedding theorems for groups," *Journal of the London Mathematical Society*, vol. 24 (1949), pp. 247-254.

[3] A. Malcev, "On the immersion of an algebraic ring into a field," *Mathematische Annalen*, vol. 113 (1937), pp. 686-691.

[4] ———, "On a representation of non-associative rings," *Uspekhi Matematicheskikh Nauk*, (N. S.), vol. 7 (1952), pp. 181-185.

[5] A. M. Turing, "The word problem in semigroups with cancellation," *Annals of Mathematics*, vol. 52 (1950), pp. 491-505.

[6] A. I. Zukov, "Reduced systems of defining relations in non-associative algebras," *Matematecheskii Sbornik*, N. S., vol. 27 (1950), pp. 267-280.

THE UNITARY EQUIVALENCE OF BINORMAL OPERATORS.*[1]

By ARLEN BROWN.

1. Introduction. The principal object of this paper is to give a set of unitary invariants for a certain class of operators (called binormal) so chosen as to be a generalization in operator form of the 2×2 complex valued matrix in a way parallel to the way in which a normal operator is a generalization of a complex number. A binormal operator consists essentially of a 2×2 matrix of mutually commuting normal operators acting in the usual fashion on a Hilbert space of the form $\mathcal{H} \oplus \mathcal{H}$. But while this characterization of the object of investigation has the advantage of immediacy, the more convenient definition for the purposes of this paper, and the one given below, is in terms of the *-ring it generates. The exact relation between these two aspects of a binormal operator is the first problem to which we address our attention, and the desired formulation is stated in Theorems 1 and 2. The central result of the paper may be paraphrased as follows: We attach to a binormal operator four commuting normal operators (fixed polynomial functions of the operator and its adjoint) and show that except for comparatively uninteresting normal direct summands which each posseses, two binormal operators are unitarily equivalent if and only if these four functions of each are simultaneously unitarily equivalent in pairs (Theorem 5). In the course of proving this it is necessary to obtain considerable information about the structure of a binormal operator, and in particular a sort of standard form is given for any such operator (§ 9). Also we show that Dixmier's solution of the "two projection problem" [2] is subsumed under the theory of binormal operators in a natural and satisfactory manner.

Concepts or terms introduced below without any definition are to be taken as defined in such standard works as [11] or [15]. All the terminology will be found to be in accord with the more recent work [4]. To mention a few relevant points in advance however: Hilbert space always means a Hilbert

* Received Nov. 1953.

[1] This paper is essentially a part of the thesis submitted in partial fulfillment of the requirements for the Ph. D. degree at The University of Chicago. The research for the paper was carried out while the author was an Atomic Energy Commission Pre-Doctoral Fellow.

space admitting complex numbers as scalars. It is not assumed to be infinite dimensional. The term operator is used exclusively to signify a bounded linear transformation of a Hilbert space into itself. The notions of a ring (of operators), of the equivalence of projections, of factors and types etc., are those introduced by Murray and von Neumann in [10]. Rings are not assumed to contain the identity operator (but every ring does automatically have a unit, namely its maximal central projection).

The familiar notation $\mathrm{Diag}(x, y, z, \cdots)$ ($x, y, z, \cdots$ any symbols) is employed to signify the diagonal matrix with entries $x, y, z, \cdots$ from left to right down the main diagonal. Moreover, for the sake of brevity in repeated references to triangular 2×2 matrices we use the notation $\mathrm{Tri}(x, y, z)$ to stand for

$$\begin{pmatrix} x & y \\ 0 & z \end{pmatrix}.$$

The author wishes to express his sincere gratitude to Professor Irving Kaplansky for his helpful suggestions during the preparation of this paper. Thanks are particularly due for many conversations about the papers on AW^*-algebras [6], [8] which are the basis of this study.

2. Binormal Rings. We begin by defining the class of operators to be investigated in terms of the rings containing them. A ring $\boldsymbol{R}$ will be called *binormal* if it satisfies the identity

$$(*) \qquad \sum \pm X_{i_1} X_{i_2} X_{i_3} X_{i_4} = 0$$

where X_i, $i = 1, \cdots, 4$ are arbitrary elements of $\boldsymbol{R}$, the summation is taken over all permutations of $(1, 2, 3, 4)$ and the sign is the parity of the permutation. A single operator is *binormal* if the ring it generates (the smallest ring containing it) is binormal. This identity is so chosen that the full ring of $n \times n$ matrices over a commutative ring with unit satisfies it if and only if $n \leqq 2$ [1], and it would not be very difficult to show directly that the only binormal factors are I_1 and I_2, thus making it possible to use reduction theory to study the relation between binormal rings and matrix rings, and indeed the next two sections will make precise the exact sense in which a binormal ring is a 2×2 matrix ring. It has seemed preferable, however, to develop the necessary theory "in the large" independently of the use of reduction theory, and the following analysis is based upon some results of Kaplansky [8].

3. The imposition of (*) on a ring of operators restricts the equivalence of projections in the ring. Thus we have at once:

LEMMA 3.1. *If the ring of operators* $\boldsymbol{R}$ *is binormal then it does not contain three mutually equivalent, orthogonal, nonzero projections.*

Proof. If it did one could construct at once in standard fashion an abstract system of 3×3 matrix units, which obviously could not satisfy (*).

Let us follow the terminology of [8] in calling a projection $E \neq 0$ in a ring $\boldsymbol{R}$ abelian if the subring $E\boldsymbol{R}E$ is abelian. The following lemma is an immediate consequence of Lemma 3.1 and [8], Lemma 4.12.

LEMMA 3.2. *If* $\boldsymbol{R}$ *is a nontrivial binormal ring then* $\boldsymbol{R}$ *contains an abelian projection. In fact, if* $E \neq 0$ *is any projection in the center of* $\boldsymbol{R}$ *then* $\boldsymbol{R}$ *contains an abelian projection* $F \leqq E$. (*In terminology of* [8] *then,* $\boldsymbol{R}$ *is of type I.*)

We call a projection G in a ring $\boldsymbol{R}$ *homogeneous* (*of order two*) if $G = G_1 + G_2$ where G_1 and G_2 are orthogonal and equivalent abelian projections in $\boldsymbol{R}$. (There is no need here to define homogeneity of higher orders as we shall be concerned exclusively with binormal rings.) A binormal ring will be called *homogeneous* if the identity operator 1 is homogeneous in it. The following lemma is a direct consequence of the definition.

LEMMA 3.3. *The supremum of an orthogonal family of central abelian* [*homogeneous*] *projections is itself central abelian* [*homogeneous*].

THEOREM 1. *Let* $\mathfrak{H}$ *be a Hilbert space and* $\boldsymbol{R}$ *a binormal ring on* $\mathfrak{H}$. *Then there are subspaces* $\mathfrak{M}$ *and* $\mathfrak{N}$, *orthogonal complements of one another, and rings* $\boldsymbol{R}_1$ *and* $\boldsymbol{R}_2$ *on* $\mathfrak{M}$ *and* $\mathfrak{N}$ *respectively, such that* $\boldsymbol{R}_1$ *is abelian,* $\boldsymbol{R}_2$ *is binormal and homogeneous, and such that* $\boldsymbol{R} = \boldsymbol{R}_1 \oplus \boldsymbol{R}_2$ (*i. e., such that* $\boldsymbol{R}$ *consists of all operators on* $\mathfrak{H} = \mathfrak{M} \oplus \mathfrak{N}$ *of the form* $A \oplus B$ *with* $A \in \boldsymbol{R}_1$ *and* $B \in \boldsymbol{R}_2$).

Proof. By Zorn's Lemma let $\{E_i\}$ be an orthogonal system of nonzero central projections in $\boldsymbol{R}$ which is maximal with respect to the property that each E_i is either abelian or homogeneous. It results in standard fashion from an application of Lemmas 3.1 and 3.2 along with [8], Lemma 4.8 and the usual argument from maximality that $E = \sup_i E_i$ is the unit of $\boldsymbol{R}$. We now simply split $\{E_i\}$ into two parts, denoting by P the supremum of the subcollection of the E_i's which are abelian and by Q the supremum of the homogeneous E_i's. Then $E = P + Q$, P is central abelian and Q is central and

homogeneous. We set $\mathfrak{N} = Q(\mathfrak{H})$ and $\mathfrak{M} = \mathfrak{H} \ominus \mathfrak{N}$ and define $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ by restricting the operators of R to $\mathfrak{M}$ and $\mathfrak{N}$ respectively. The desired relations follow at once.

In view of this theorem we may limit our considerations to binormal rings which are homogeneous. Accordingly, let $\boldsymbol{R}$ be a binormal ring on a Hilbert space $\mathfrak{H}$ and let $1 = E_1 + E_2$ where the projections E_i $(i = 1, 2)$ are orthogonal, equivalent and abelian in $\boldsymbol{R}$. Set $E_i(\mathfrak{H}) = \mathfrak{N}_i$. Since the E_i's are equivalent there is a partial isometry $V \varepsilon \boldsymbol{R}$ with $V^*V = E_1$ and $VV^* = E_2$, and hence $V|\mathfrak{N}_1$ is an isomorphism between the two Hilbert spaces $\mathfrak{N}_1$ and $\mathfrak{N}_2$. If we let $\mathfrak{K}$ denote the Hilbert space $\mathfrak{N}_1 \oplus \mathfrak{N}_1$ and define $j(x, y) = x + Vy$ for $(x, y) \varepsilon \mathfrak{K}$, it is easy to see that j is an isomorphism of $\mathfrak{K}$ onto $\mathfrak{H}$, and hence that the set $\hat{\boldsymbol{R}}$ of all operators $j^{-1}Aj$ for $A \varepsilon \boldsymbol{R}$ is a unitarily isomorphic copy of $\boldsymbol{R}$ on $\mathfrak{K}$. Thus we represent $\boldsymbol{R}$ as a ring of 2×2 matrices in the usual fashion. (Quite precisely, we identify the operator C on $\mathfrak{K}$ with the matrix (W_{ij}) according to the formula $C(x, y) = (W_{11}x + W_{12}y, W_{21}x + W_{22}y)$ where, of course, the W_{ij}'s are operators on $\mathfrak{N}_1$.)

It is easy to settle the question of exactly what matrices correspond to $\boldsymbol{R}$ in this scheme. In the first place, the two projections E_1 and E_2, which we now prefer to call E_{11} and E_{22}, along with V and V^*, which we denote by E_{21} and E_{12} respectively, give us a set of matrix units in $\boldsymbol{R}$ with $E_{11} + E_{22} = 1$ and, furthermore, these four operators are carried by j onto the unit matrices

$$(**) \qquad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

in natural order. Let us state the following theorem, making use of the notation already introduced.

THEOREM 2. *Let $\boldsymbol{R}$ be a homogeneous binormal ring on $\mathfrak{H}$ with the matrix units $\{E_{ij}\}$ $i, j = 1, 2$ where E_{11} and E_{22} are orthogonal abelian projections with $E_{11} + E_{22} = 1$. Then $\boldsymbol{R}$ is represented via j on $\mathfrak{K}$ as the ring $\hat{\boldsymbol{R}}$ of exactly those matrices (W_{ij}) whose entries lie in the abelian ring $\boldsymbol{Z}'$ on $\mathfrak{N}_1$, where $\boldsymbol{Z}'$ consists of all restrictions to $\mathfrak{N}_1$ of the operators in $E_{11}\boldsymbol{R}E_{11}$. A ring is homogeneous and binormal if and only if it can be so represented for some abelian ring containing 1. Under this isomorphism the matrix units E_{ij} correspond in natural order to the unit matrices (**) and the center is carried onto the 2-fold copy of $\boldsymbol{Z}'$ consisting of all matrices* $\mathrm{Diag}(W, W)$ *with $W \varepsilon \boldsymbol{Z}'$.*

Proof. In the first place, it is a well known fact of matrix theory (see

e. g. [5], pp. 57-58) that $\boldsymbol{R}$ consists of all elements of the form $\Sigma_{ij}A_{ij}E_{ij}$ where $A_{ij} \varepsilon \boldsymbol{D}$ and $\boldsymbol{D}$ denotes the collection of elements of $\boldsymbol{R}$ commuting with all E_{ij}. This $\boldsymbol{D}$ is a ring between $\boldsymbol{R}$ and its center $\boldsymbol{Z}$ and an operator in it is carried by j onto some matrix (W_{ij}) which commutes with all four unit matrices. Computation shows that $W_{12} = W_{21} = 0$ and $W_{11} = W_{22}$ so that the elements of $\boldsymbol{D}$ correspond under j to matrices $\mathrm{Diag}(W, W)$ in $\boldsymbol{R}$. But then also it is clear from the definition that $W \varepsilon \boldsymbol{Z}'$ and hence that $\boldsymbol{D}$ is abelian, so that $\boldsymbol{D} = \boldsymbol{Z}$ is the center of $\boldsymbol{R}$, and, as is readily seen, the theorem follows.

COROLLARY A. *In a homogeneous binormal ring, the only central abelian projection is* 0.

Proof. A central projection must have the form $\mathrm{Diag}(E, E)$ where E is a projection. If this element is also abelian the matrix $\mathrm{Tri}(0, E, 0)$ must commute with its own transpose, whence $E = 0$.

This last result enables us to clear up a point which was left unsettled in Theorem 1. Reverting for the time being to the notation of that theorem we recall that the unit E of the arbitrary binormal ring $\boldsymbol{R}$ is written $E = P + Q$ with $\mathfrak{N} = Q(\mathfrak{H})$ so that the result, $\boldsymbol{R}_2$, of restricting the elements of $\boldsymbol{R}$ to $\mathfrak{N}$ is homogeneous and binormal. Now if F is any central abelian projection in $\boldsymbol{R}$ and if $F_2 = F|\mathfrak{N}$ then F_2, being again central and abelian in $\boldsymbol{R}_2$, must be trivial, or $FQ = 0$ which implies $F \leqq P$. Since P is itself central abelian by construction we see that P is identified as the maximal central abelian projection in $\boldsymbol{R}$.

COROLLARY B. *The decomposition of Theorem* 1 *is unique. The projection on the subspace* $\mathfrak{M}$ *is the sum of the annihilating projection* $1 - E$ *and the maximal central abelian projection* P.

4. If $\mathfrak{H}_1$ and $\mathfrak{H}_2$ are Hilbert space and if $\phi: \mathfrak{H}_1 \to \mathfrak{H}_2$ is an isomorphism then we write $\phi^0(A) = \phi^{-1}A\phi$ where A is an operator on $\mathfrak{H}_2$, introducing the notation ϕ^0 for the unitary isomorphism implemented by ϕ. We proceed to investigate the circumstances under which two binormal rings are unitarily equivalent, and in view of the last result we shall assume that they are both homogeneous.

THEOREM 3. *Let* $\boldsymbol{R}_1$ *and* $\boldsymbol{R}_2$ *be homogeneous binormal rings on Hilbert spaces* $\mathfrak{H}_1$ *and* $\mathfrak{H}_2$ *respectively. Let* $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2$ *be their respective centers. Then* $\boldsymbol{R}_1$ *and* $\boldsymbol{R}_2$ *are unitarily equivalent if and only if* $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2$ *are, and*

further, if ϕ^0 is any unitary isomorphism of $\mathbf{Z}_2$ onto $\mathbf{Z}_1$, there is another unitary isomorphism ψ^0 of $\mathbf{R}_2$ onto $\mathbf{R}_1$ which agrees with ϕ^0 on the ring $\mathbf{Z}_2$.

Proof. We limit our attention to the last assertion. By Theorem 2 we may assume that $(i = 1, 2)$ $\mathcal{H}_i = \mathcal{M}_i \oplus \mathcal{M}_i$ and that the operators in the ring $\mathbf{R}_i$ are those matrices whose entries lie in the abelian ring $\mathbf{Z}'_i$ on the space $\mathcal{M}_i$. The centers, $\mathbf{Z}_i$, consist then of all operators $A \oplus A$ with $A \in \mathbf{Z}'_i$. It is sufficient to obtain a new isomorphism $\psi: \mathcal{H}_1 \to \mathcal{H}_2$ such that $\psi^0|\mathbf{Z}_2 = \phi^0|\mathbf{Z}_2$ and with the further property that ψ can be factored, $\psi = \omega \oplus \omega$, where $\omega: \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism with $\omega^0(\mathbf{Z}'_2) = \mathbf{Z}'_1$. Moreover, if each $\mathbf{Z}'_i$ (and hence each $\mathbf{Z}_i$) is resolved into its pieces of uniform multiplicity and account is taken of the given unitary isomorphism, it follows that it suffices to prove the theorem under the assumption that the rings $\mathbf{Z}'_i$ have uniform multiplicity n (where n is a not necessarily finite cardinal number) and we accordingly assume this to be the case.[2]

Now $(i = 1, 2)$ $\mathbf{Z}'_i$ is the n-fold copy of a maximal abelian ring $\mathbf{M}_i$ on some Hilbert space $\mathcal{K}_i$ (for the sake of simplicity we ignore the distinction between internal and external direct sum) and the given isomorphism ϕ furnishes us with an algebraic *-isomorphism $\chi^0: \mathbf{M}_2 \to \mathbf{M}_1$ in a natural way. But these are maximal abelian rings and a *-isomorphism between two such rings is in fact spatially implemented.[3] Hence there exists an isomorphism $\chi: \mathcal{K}_1 \to \mathcal{K}_2$ which justifies our use of the notation χ^0. The desired isomorphisms ω and ψ are obtained as n-fold and $2n$-fold copies respectively of χ.

5. The most general homogeneous binormal ring is the ring $\mathbf{R}$ of operators on a space $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}$ consisting of all matrices with entries in $\mathbf{Z}'$, where $\mathbf{Z}'$ is some abelian ring on $\mathcal{M}$ containing 1. Concerning the ring $\mathbf{Z}'$ it is well known that there is a compact Hausdorff space $\mathcal{X}$ such that $\mathbf{Z}'$ is isomorphic to the algebra $C(\mathcal{X})$ of all complex valued continuous functions on $\mathcal{X}$ under a map which preserves not only the ring operations in $\mathbf{Z}'$ considered as a *-ring but the norms of elements as well. This is no more than the usual representation theorem for abelian C^*-algebras with unit, and

[2] It may be noted that from this point the proof goes through even if the *-isomorphism ϕ^0 is not spatially implemented. Thus the assumption that the isomorphism is unitary serves essentially to align the various parts of the rings $\mathbf{Z}_i$ having uniform multiplicity. It is not difficult to see that the condition of uniform multiplicity here imposed is equivalent to the homogeneity of the commutants in an appropriate sense. See [6].

[3] A proof of this important fact has been given by Segal in [13], Part II, Th. 1 and [14], Th. 4. 1.

would hold if $\boldsymbol{Z}'$ were only closed in the uniform topology for operators. The fact that $\boldsymbol{Z}'$ is actually a ring of operators leads to the additional requirement that $\mathfrak{X}$ be totally disconnected (i. e., possess a topological base of compact-open sets) and further to the condition that the closure of every open set in $\mathfrak{X}$ be open ([16]).

We conclude our considerations of binormal rings by extending this correspondence to a functional representation for $\boldsymbol{R}$ itself. Consider the collection $M_2(\mathfrak{X})$ of all continuous functions from $\mathfrak{X}$ to the ring of all 2×2 matrices. (The latter ring is, for this purpose, considered as the ring of all operators on some 2-dimensional Hilbert space. The topology that it so acquires is obviously the same as that commonly given it by means of coordinates.) We can give $M_2(\mathfrak{X})$ the structure of a normed algebra in the usual way by defining algebraic operations pointwise and writing $\| A \| = \sup_x \| A(x) \|$ for $A \in M_2(\mathfrak{X})$. It is a simple matter to verify that this yields a new C^*-algebra which we also denote by $M_2(\mathfrak{X})$.

Now an element W of $\boldsymbol{R}$ is a matrix with entries, say, W_{ij} for $i, j = 1, 2$. The correspondence between $\boldsymbol{Z}'$ and $C(\mathfrak{X})$ gives us continuous functions w_{ij} on $\mathfrak{X}$ corresponding respectively to W_{ij}. We define $\Phi(W)(x) = (w_{ij}(x))$ for $x \in \mathfrak{X}$. Then $\Phi(W) \in M_2(\mathfrak{X})$ and we have a mapping $\Phi: \boldsymbol{R} \to M_2(\mathfrak{X})$. One verifies directly that Φ is onto $M_2(\mathfrak{X})$ and preserves all the ring operations in $\boldsymbol{R}$ (including *) and, from the uniqueness of the norm of a C^*-algebra ([7], Th. 6. 4), we conclude also that Φ preserves the norms of elements. For convenience we summarize this as:

THEOREM 4. *Let $\boldsymbol{R}$ be the homogeneous binormal ring of operators on $\mathfrak{H} = \mathfrak{M} \oplus \mathfrak{M}$ consisting of matrices with entries in the abelian ring $\boldsymbol{Z}'$ on $\mathfrak{M}$. Let $C(\mathfrak{X})$ be isomorphic to $\boldsymbol{Z}'$ under a map ϕ. Then the map Φ defined by $\Phi(W_{ij})(x) = (\phi(W_{ij})(x)$ is an isometric *-isomorphism (or in other words a C^*-isomorphism) between $\boldsymbol{R}$ and the ring $M_2(\mathfrak{X})$.*

6. Binormal operators. Call a subspace $\mathfrak{L}$ *normal* with respect to A if it reduces A and if $A|\mathfrak{L}$ is a normal operator. The supremum $\mathfrak{K}$ of all subspaces normal with respect to A is also normal with respect to A. This maximum normal subspace, together with the normal operator $A|\mathfrak{K}$ on it, we shall call the *normal kernel*[4] of A. If $\mathfrak{K} = 0$ we shall say that A is *reduced*. In terms of these concepts we may characterize the binormal operators which generate homogeneous rings.

[4] We shall sometimes take the liberty of referring to either the subspace $\mathfrak{K}$ or the operator $A|\mathfrak{K}$ alone as the normal kernel. The exact meaning will always be clear in context.

LEMMA 6.1. *Let A be binormal on $\mathfrak{H}$ and let $\mathfrak{K}$ be its normal kernel. Let A generate the binormal ring $\mathbf{R}$ and let $\mathfrak{M}$ and $\mathfrak{N}$ be the subspaces of the unique decomposition of Theorem* 1. *Then $\mathfrak{K} = \mathfrak{M}$.*

Proof. In the first place, $\mathfrak{M}$ is normal with respect to A by its definition so that $\mathfrak{M} \subset \mathfrak{K}$. To prove the other inclusion, let T be any operator commuting with A and A^*. Computation shows that $\mathfrak{K}$ is invariant under T. Replacing T by T^* we see that $\mathfrak{K}$ reduces T. Write F for the projection of $\mathfrak{H}$ onto $\mathfrak{K}$ and E as before for the maximal projection in $\mathbf{R}$. Then ([12], *Satz* 5) $FE \in \mathbf{R}$. But FE is also of course in the commutant of $\mathbf{R}$ and hence is central. Let $\mathfrak{L} = E(\mathfrak{H})$ so that FE projects onto $\mathfrak{K} \cap \mathfrak{L}$. It is easily seen that the restriction of any two operators in $\mathbf{R}$ to $\mathfrak{K} \cap \mathfrak{L}$ must commute, and hence that FE is abelian. By Theorem 2, Corollary B, this implies that $\mathfrak{K} \cap \mathfrak{L} \subset \mathfrak{M}$ and the desired relation $\mathfrak{K} \subset \mathfrak{M}$ follows.

In other words then, the binormal operators which generate homogeneous rings are those which are reduced. Since the normal kernel is surely preserved under unitary isomorphism, we may accomplish our goal of giving a set of unitary invariants for a binormal operator by solving the problem in the reduced case.

7. Before turning to the general investigation it will be instructive to look closely at two special kinds of binormal operators which are particularly amenable to our approach.

LEMMA 7.1. *Let T_1 and T_2 be operators on $\mathfrak{H}$ satisfying*

$$T_i^2 = \alpha_i T_i \qquad i = 1, 2$$

for complex numbers α_1 and α_2. Then the algebra of operators generated by T_1 and T_2 (in the purely algebraic sense) satisfies (*).

Proof. This is a special statement of the result ([9], Lemma 5).

In view of this it is easily seen that an operator A is certainly binormal if $A^2 = \alpha A$ for some complex number α, for then also $A^{*2} = \bar{\alpha} A^*$. Indeed, it is not hard to see that A is binormal if it satisfies a quadratic equation $\alpha A^2 + \beta A + \gamma = 0$ for α, β and γ arbitrary complex numbers, but all the possibilities thus presented reduce readily to one or the other of the following distinct special cases, which we proceed to treat in order.

a) $A^2 = 0$: The assumption that A is reduced leads in this case to $\mathfrak{N} \cap \mathfrak{N}^* = 0$ where $\mathfrak{N}$ and $\mathfrak{N}^*$ are the null spaces of A and A^* respectively.

Let $\boldsymbol{R}$ be the ring generated by such an A. Then (employing the notation introduced in Theorem 2) $\boldsymbol{R}$ consists of all 2×2 matrices (Z_{ij}) with entries in the ring $\boldsymbol{Z}'$ on $\mathfrak{M}$. The operator A itself we shall write as the matrix (A_{ij}). Using Theorem 4, we obtain an isomorphism between $\boldsymbol{R}$ and a ring $M_2(\mathfrak{X})$ under which A corresponds to an element $A(\cdot)$ whose value at each point $x \in \mathfrak{X}$ is the matrix $A(x) = (a_{ij}(x))$ with $A(x)^2 = 0$. (Here and throughout, if $Z \in \boldsymbol{Z}'$ we shall denote by z the continuous function in $C(\mathfrak{X})$ corresponding to it, and if $B = (Z_{ij})$ is in $\boldsymbol{R}$, the corresponding element of $M_2(\mathfrak{X})$ will be denoted by $B(\cdot)$ so that $B(x) = (z_{ij}(x))$ for $x \in \mathfrak{X}$.) For every x the trace and determinant of $A(x)$ are both 0. Now set

$$C_1 = \begin{pmatrix} A_{11}^* & A_{21}^* \\ A_{21} & -A_{11} \end{pmatrix} \qquad C_2 = \begin{pmatrix} A_{12}^* & A_{22}^* \\ A_{22} & -A_{12} \end{pmatrix}.$$

Since the operators $C_1C_1^* = C_1^*C_1$ and $C_2C_2^* = C_2^*C_2$ are central in $\boldsymbol{R}$, the projections, says F_1 and F_2, on their respective null spaces are also central, and correspond to compact-open sets L_1 and L_2 in $\mathfrak{X}$. (More exactly, F_i corresponds to the function in $M_2(\mathfrak{X})$ whose value is the matrix 1 at every point of L_i and 0 at every point of $\mathfrak{X} - L_i$.) The subspaces onto which F_1 and F_2 project (the null spaces of C_1 and C_2 respectively) we shall call $\mathfrak{L}_1$ and $\mathfrak{L}_2$. Write C_i in polar form, $C_i = |C_i|U_i$ for $i = 1, 2$. (The absolute value of a normal operator is the positive square root of the product of it with its own adjoint.) The operator U_i is 0 on the subspace $\mathfrak{L}_i$, (which reduces it) and operates unitarily on $\mathfrak{H} \ominus \mathfrak{L}_i$ and so corresponds to a function $U_i(\cdot) \in M_2(\mathfrak{X})$ which is 0 on L_i and unitary valued on $\mathfrak{X} - L_i$.

Now define

$$U(x) = \begin{cases} U_1(x) & x \notin L_1 \\ U_2(x) & x \in L_1 \end{cases} \qquad C(x) = \begin{cases} C_1(x) & x \notin L_1 \\ C_2(x) & x \in L_1. \end{cases}$$

These correspond to the operators $U = U_1 + F_1U_2$ and $C = C_1 + F_1C_2$ in $\boldsymbol{R}$ and $C = |C|U$ is the polar resolution of C. The function $U(\cdot)$ is unitary valued except on $L_1 \cap L_2$. But on this set, $C_1(x) = C_2(x) = 0$ so that $a_{ij}(x) = 0$ for $i, j = 1, 2$ and $A(x) = 0$. This means that $F_1F_2A = 0$ and that $\mathfrak{L}_1 \cap \mathfrak{L}_2$ is normal with respect to A. Since A is reduced we conclude that $L_1 \cap L_2 = 0$, or, in other words, that U is unitary. In particular, C is non-singular.[5]

Now denote by B the operator UAU^*. We have $CC^*B = CAC^*$ and it follows that CC^*B has the form $\mathrm{Tri}(0, *, 0)$ since $C(x)A(x)C(x)^*$ has this form whether $x \in L_1$ or not by the way C_1 and C_2 were chosen. But then also

[5] A non-singular operator is one with a trivial null space.

$B = \mathrm{Tri}(0, G, 0)$ for some $G \in \boldsymbol{Z}'$, since CC^* is a non-singular matrix $\mathrm{Diag}(P, P)$, which implies that P can be cancelled in the matric equation for CC^*B.

Since B must be reduced also we learn that G is non-singular. Hence, in the polar resolution $G = |G|W$, W is unitary and the operator $U' = \mathrm{Diag}(1, W)$ is unitary in $\boldsymbol{R}$. But $U'BU'^* = \mathrm{Tri}(0, |G|, 0)$. Thus by a slight modification in the unitary operator U we can arrange for B to have the property that $G > 0$, and we accordingly assume this to be the case.

Now polynomials in B and B^* are matrices whose entries are polynomials in G. A continuity argument shows that every element of the ring generated by B, which is of course $\boldsymbol{R}$ itself, has its entries in the ring on $\mathfrak{M}$ generated by G. This in turn implies that G generates $\boldsymbol{Z}'$, and hence that

$$B^*B + BB^* = \begin{pmatrix} G^2 & 0 \\ 0 & G^2 \end{pmatrix}$$

generates the center $\boldsymbol{Z}$ of $\boldsymbol{R}$. This central element may also, of course, be written $A^*A + AA^*$. Thus $A^*A + AA^*$ generates the center $\boldsymbol{Z}$ of the ring $\boldsymbol{R}$ generated by A.

Now suppose given reduced operators A_1 and A_2 on spaces $\mathfrak{H}_1$ and $\mathfrak{H}_2$ respectively, such that both A_1 and A_2 have square zero. Suppose moreover that there exists an isomorphism $\phi\colon \mathfrak{H}_1 \to \mathfrak{H}_2$ with

$$\phi^0(A_2^*A_2 + A_2A_2^*) = A_1^*A_1 + A_1A_1^*.$$

Let A_i generate the ring of operators $\boldsymbol{R}_i$ with center $\boldsymbol{Z}_i$, $i = 1, 2$. Then $\phi^0(\boldsymbol{Z}_2) = \boldsymbol{Z}_1$, and, emplyong Theorem 3, we may replace ϕ by an isomorphism ψ such that not only is $\psi^0(A_2^*A_2 + A_2A_2^*) = A_1^*A_1 + A_1A_1^*$ and $\psi^0(\boldsymbol{Z}_2) = \boldsymbol{Z}_1$, but also $\psi^0(\boldsymbol{R}_2) = \boldsymbol{R}_1$. We now make $\boldsymbol{R}_1$ correspond to a ring $M_2(\mathfrak{X})$ as above. As we have seen, there are unitary elements U_i $(i = 1, 2)$ in $\boldsymbol{R}_1$ such that $U_1A_1U_1^*$ and $U_2\psi^0(A_2)U_2^*$ correspond to elements $G_i(\cdot)$ in $M_2(\mathfrak{X})$ with $G_i(x) = \mathrm{Tri}(0, g_i(x), 0)$, $g_i(x) \geqq 0$ for $x \in \mathfrak{X}$. Since also

$$\mathrm{Diag}(g_1^2(x), g_1^2(x)) = \mathrm{Diag}(g_2^2(x), g_2^2(x)),$$

we have $g_1(x) = g_2(x)$ for every x so that $G_1(\cdot) = G_2(\cdot)$. Thus $U_1A_1U_1^* = U_2\psi^0(A_2)U_2^*$. The appropriate combination of the isomorphisms U_1, U_2 and ψ^0 yields a unitary equivalence between A_1 and A_2.

A simple argument frees us from the assumption that A is reduced and we summarize as follows: *For an operator A with $A^2 = 0$ a complete set of unitary invariants is furnished by the unitary equivalence class of the self-adjoint operator $A^*A + AA^*$.* It may be remarked that in the reduced case

the two operators A^*A and AA^* are themselves unitarily equivalent so that we could go a step further and use either of them in determining an invariant for $A^*A + AA^*$ and hence for A. It is easy to see, however, that this assertion fails in the presence of a non-trivial normal kernel

b) $A^2 = A$: We assume once more that A is without normal kernel. The latter in this case is clearly a projection. The ring $\boldsymbol{R}$ generated by A may be represented on $M_2(\mathfrak{X})$ under a correspondence which carries A onto a function $A(\cdot)$ with $A(x)^2 = A(x)$ everywhere on $\mathfrak{X}$. The matrix $A(x)$ can have only 0's and 1's for characteristic roots so that its trace $t(x)$ is either 0, 1 or 2. Since t is continuous the subset on which t takes each of these values is compact-open. But let $d(x)$ denote the determinant of $A(x)$. Then $A(x)^2 - t(x)A(x) + d(x) = (1 - t(x))A(x) + d(x) = 0$, so that the possibilities $t(x) = 0$ and $t(x) = 2$ yield respectively the relations $A(x) = \mp d(x)$, and the sets on which these two relations hold correspond to subspaces which are normal with respect to A. It follows that $t(x) = 1$ identically and hence that d vanishes identically. Let us again set $B = UAU^*$ where U is the same unitary element defined in a). In the present case it is readily seen that the element of $M_2(\mathfrak{X})$ corresponding to $CC^*B = CAC^*$ has the form

$$\begin{pmatrix} p(x) & q(x) \\ 0 & 0 \end{pmatrix} \qquad x \in \mathfrak{X}$$

where, as before, $CC^* = \mathrm{Diag}(P, P)$ and it follows that $B = \mathrm{Tri}(1, H, 0)$ for some operator $H \in \boldsymbol{Z}'$. As before we may assume that $H > 0$ since it is non-singular.

By a repetition of the argument used above we conclude that the ring $\boldsymbol{Z}'$ on $\mathfrak{M}$ is generated by the single operator H (1 being in this latter ring automatically). Since in this case

$$(B - B^*)^2 = \begin{pmatrix} -H^2 & 0 \\ 0 & -H^2 \end{pmatrix}$$

$(B - B^*)^2 = (A - A^*)^2$ generates the center $\boldsymbol{Z}$ of $\boldsymbol{R}$, and, proceeding exactly as before, we find: *A complete set of unitary invariants for the idempotent operator A is furnished by the unitary equivalence class of the self-adjoint operator* $(A - A^*)^2$ *(or of* $A - A^*$*) together with that of the projection which is the normal kernel of A.*

8. In these two examples the procedure has been to put the information we have about binormal rings and what we know of 2×2 matrices together. In approaching the general case in this spirit we shall need a solution of the special problem of the unitary equivalence of 2×2 matrices.

We begin by assuming that the matrix C is not normal. Let b be a characteristic root of C. It is a simple matter to find a unitary matrix U such that UCU^* has the form $\mathrm{Tri}(a, c, b)$ with $(0, b)$ in the lower row. The entry a is then the other characteristic root and is fixed, but with these three entries unchanged the matrix U may be varied so as to replace c by $ce^{i\phi}$ with arbitrary ϕ. In particular, c can be made real and positive. With this restriction, along with the condition that the lower row be $(0, b)$, the matrix equivalent to C is uniquely determined. Similarly C is equivalent to another matrix $\mathrm{Tri}(b, d, a)$ with $d > 0$, but computation shows that $|c| = |d|$, or $c = d$. This proves:

LEMMA 8.1. *If C is any 2×2 complex matrix then there is a unique non-negative real number r such that C is unitarily equivalent to a matrix $\mathrm{Tri}(a, r, b)$. The diagonal entries are then the characteristic roots of C in some order (so that while this is not quite a canonical form, there can be at most two choices for it). A complete set of unitary invariants for C is furnished by the number r and the trace and determinant of C.*

Let us now define (for A an element of an arbitrary C^*-algebra) the functions:

$$f_1(A) = [A, A^*]^2 \quad f_2(A) = [A^2, A^*]^2, \quad f_3(A) = ([A, A^*] + [A^2, A^*])^2,$$
$$f_4(A) = (A[A, A^*])^2,$$

where $[A, B]$ denotes the commutator $BA - AB$.

LEMMA 8.2. *If A is a 2×2 complex matrix then $f_i(A)$ is a scalar matrix for $i = 1, \cdots, 4$. Two non-normal 2×2 matrices A and B are unitarily equivalent if and only if $f_i(A) = f_i(B)$ for all i.*

Proof. Since the square of a 2×2 matrix with zero trace, and hence in particular of a commutator, is scalar, the first part of the assertion is readily established. It follows that the functions f_i are constant on unitary equivalence classes, which proves the necessity of the condition of the second part. Also, to show sufficiency, we may replace A and B by any matrices unitarily equivalent to them, assuming for instance that $A = \mathrm{Tri}(a_1, r_1, b_1)$ and $B = \mathrm{Tri}(a_2, r_2, b_2)$. Computation shows that

$$f_2(A) = t_1^2 f_1(A), \qquad f_3(A) = (t_1 + 1)^2 f_1(A), \qquad f_4(A) = d_1 f_1(A)$$

where t_1 and d_1 are the trace and determinant of A and where $f_1(A) = r_1^2(r_1^2 + |a_1 - b_1|^2)$. The analogous relations hold for B.

Now since we assume explicitly that A and B are not normal $f_1(A) = f_1(B) \neq 0$, and we may divide the other qualities by this one, obtaining the relations $t_1 = t_2$ and $d_1 = d_2$. This means that A and B have the same characteristic roots, and, in particular, that $|a_1 - b_1|^2 = |a_2 - b_2|^2$. From this and from $f_1(A) = f_1(B)$ it follows that $r_1 = r_2$, and the lemma is proved.

9. Before stating the final theorem we have need of one more preliminary result.

LEMMA 9.1. *If B and C are any elements of an abelian ring of operators, then there is an operator X in the ring satisfying $X^2 + BX + C = 0$.*

Proof. In the special case $B = 0$ this amounts to finding a square root of a normal operator. In the general case we have but to complete the square as in the familiar quadratic formula.

THEOREM 5. *If A is a binormal operator then the operators $f_i(A)$ for $i = 1, \cdots, 4$ are normal and mutually commutative. If A_1 and A_2 are reduced binormal operators then they are unitarily equivalent if and only if there exists an isomorphism ϕ such that for all i $\phi^0(f_i(A_2)) = f_i(A_1)$.*

Proof. It clearly suffices to prove the first assertion also under the assumption that A has no normal kernel, while, in the second part, we again need only concern ourselves with the sufficiency of the condition. Let the reduced binormal operator A on $\mathfrak{H}$ generate the ring $\boldsymbol{R}$ with center $\boldsymbol{Z}$. Then employing Theorems 2 and 4 as before we may assume that $\mathfrak{H} = \mathfrak{M} \oplus \mathfrak{M}$, that $\boldsymbol{R}$ consists of all 2×2 matrices with entries in the abelian ring $\boldsymbol{Z}'$ on $\mathfrak{M}$, and that $\boldsymbol{R}$ is isomorphic to $M_2(\mathfrak{X})$ where $\mathfrak{X}$ is the representation space of $\boldsymbol{Z}'$. The given operator A appears as, say, the matrix (A_{ij}). Also, the operators $f_i(A)$ correspond to elements $H_i(\cdot)$ in $M_2(\mathfrak{X})$ for which $H_i(x) = f_i(A(x))$ for all x. This means that $H_i(\cdot)$ is scalar valued everywhere and hence that $f_i(A) \in \boldsymbol{Z}$ for $i = 1, \cdots, 4$. Thus the first part of the theorem is proved.

We next take up the problem of reducing the matrix (A_{ij}) to a simpler form, exactly as was done in § 7, but taking here in the general case some extra care which was not needed there. Write $T = A_{11} + A_{22}$ and $D = A_{11}A_{22} - A_{12}A_{21}$.[6] By Lemma 9.1 there is an operator $Z_2 \in \boldsymbol{Z}'$ such that $Z_2^2 - TZ_2 + D = 0$. To facilitate matters later on we shall insist

[6] These might conveniently be called the trace and determinant of A. It is seen that $T \oplus T \in \boldsymbol{Z}$ is the center valued trace of A in the sense of Dixmier [3].

further that Z_2 lie in the ring generated by T and D on $\mathfrak{M}$. This time we write

$$C_1 = \begin{pmatrix} Z_2^* - A_{11}^* & -A_{21}^* \\ A_{21} & Z_2 - A_{11} \end{pmatrix}, \qquad C_2 = \begin{pmatrix} -A_{12}^* & Z_2^* - A_{22}^* \\ Z_2 - A_{22} & A_{12} \end{pmatrix}.$$

It is easy to see that C_i is normal in $\boldsymbol{R}$ and that $C_iC_i^* \in \boldsymbol{Z}$ for $i = 1, 2$. Hence, again, if F_i projects $\mathfrak{H}$ onto the null space of C_i, F_i is central and corresponds to a compact-open set $L_i \subset \mathfrak{X}$. On L_1 we have of course $a_{21} = 0$ and $a_{11} = z_2$ while on L_2 the relations $a_{12} = 0$ and $a_{22} = z_2$ hold identically. Thus F_1F_2A is normal and $L_1 \cap L_2 = 0$. Write $C_i = |C_i|U_i$ in polar form for $i = 1, 2$. The element $U_i(\cdot) \in M_2(\mathfrak{X})$ is unitary valued on $\mathfrak{X} - L_i$ and vanishes identically on L_i. Hence if we define $U(x)$ to be $U_1(x)$ on $\mathfrak{X} - L_1$ and $U_2(x)$ on L_1, the operator $U = U_1 + F_1U_2 \in \boldsymbol{R}$ corresponding to $U(\cdot)$ is unitary. Clearly also $C = |C|U$ is the polar decomposition of C where $C = C_1 + F_1C_2$ corresponds to the function $C(\cdot)$ with $C(x) = C_1(x)$ for $x \notin L_1$ and $C(x) = C_2(x)$ for $x \in L_1$.

Now let $B = UAU^*$ and choose $P \in \boldsymbol{Z}'$ so that the central element $CC^* = C_1C_1^* + F_1C_2C_2^* = \mathrm{Diag}(P, P)$. The exact form of p may readily be found and direct computation discloses that the function corresponding to $CC^*B = CAC^*$ has the form $\mathrm{Tri}(*, *, p(x)z_2(x))$ at every point of $\mathfrak{X}$ so that $CC^*B = \mathrm{Tri}(PZ_1, PR, PZ_2)$ for certain operators Z_1 and R, whence, cancelling P once again, $B = \mathrm{Tri}(Z_1, R, Z_2)$. The operator Z_1 is clearly $T - Z_2$. Moreover, the operator R can be assumed to be positive definite as before. Thus the general reduced binormal operator A is unitarily equivalent in the ring which it generates to an operator $B = \mathrm{Tri}(Z_1, R, Z_2)$ with $R > 0$.

We shall now show that the center $\boldsymbol{Z}$ is generated by the four operators $f_i(A)$. Since these are central we have $f_i(A) = \mathrm{Diag}(M_i, M_i)$ $i = 1, \cdots, 4$ for certain $M_i \in \boldsymbol{Z}'$ and it suffices of course to show that the M_i generate $\boldsymbol{Z}'$. Also $B = \mathrm{Tri}(Z_1, R, Z_2)$ generates $\boldsymbol{R}$ too which implies that Z_1, Z_2 and R generate $\boldsymbol{Z}'$. Finally, since $f_i(A) = f_i(B)$, the M_i are, in natural order,

$$Q, \quad T^2Q, \quad (T+1)^2Q, \quad DQ$$

where $Q = R^2(R^2 + |Z_1 - Z_2|^2)$. Now Q is non-singular (if it were not we would once more be able to produce a subspace normal with respect to B), so if $\mathfrak{L}$ reduces both Q and DQ it must reduce D. In other words D is in the ring, which we may temporarily call $\boldsymbol{Y}'$, generated by the M_i. Similarly T^2 and $(T+1)^2$ are in $\boldsymbol{Y}'$ whence $T \in \boldsymbol{Y}'$. But then also $Z_1, Z_2 \in \boldsymbol{Y}'$. Finally, along with $(|Z_1 - Z_2|^2 + 2R^2)^2 = |Z_1 - Z_2|^4 + 4Q$, $\boldsymbol{Y}'$ contains the positive square root $|Z_1 - Z_2|^2 + 2R^2$ and we are able to conclude that $R \in \boldsymbol{Y}'$ and $\boldsymbol{Y}' = \boldsymbol{Z}'$.

Now let us suppose, as in the hypothesis of the theorem, that A_1 and A_2 are reduced binormal operators and that there exists an isomorphism ϕ with $\phi^0(f_i(A_2)) = f_i(A_1)$ for $i = 1, \cdots, 4$. By making a familiar application of Theorem 3 it is easy to see that we can reduce our problem to the special case which arises when A_1 and A_2 generate the same ring $\boldsymbol{R}$ and $f_i(A_1) = f_i(A_2)$. We then map $\boldsymbol{R}$ onto $M_2(\mathfrak{X})$ by Theorem 4. Denote by M the set of points at which $A_1(x)$ and $A_2(x)$ are unitarily equivalent. Denote also by $H_i(\cdot)$ the element of $M_2(\mathfrak{X})$ corresponding to $f_i(A_1) = f_i(A_2)$ and let

$$H_1(x) = \begin{pmatrix} f_1(x) & 0 \\ 0 & f_1(x) \end{pmatrix}.$$

We know that $A_1(x)$ and $A_2(x)$ are normal if and only if $f_1(x) = 0$. Since $f_i(A_1(x)) = f_i(A_2(x))$ for $i = 1, \cdots, 4$ and for all x, it follows from Lemma 8.2 that if $f_1(x) \neq 0$ then $x \in M$. But the operators A_i are reduced and therefore the functions $A_i(\cdot)$ cannot be normal on an open subset of $\mathfrak{X}$, whence it follows that M is dense.

Now if $t_i(x)$ and $d_i(x)$ are the trace and determinant respectively of $A_i(x)$ then $t_1(x) = t_2(x)$ and $d_1(x) = d_2(x)$ for $x \in M$. Since M is dense $t_1 = t_2$ and $d_1 = d_2$ and therefore the pair of operators T and D above is the same whether constructed from A_1 or A_2. Hence for a suitably chosen Z_2 satisfying $Z_2^2 - TZ_2 + D = 0$ and for $Z_1 + Z_2 = T$ we see that A_1 and A_2 are unitarily equivalent in $\boldsymbol{R}$ to operators B_1 and B_2 respectively with $B_1 = \mathrm{Tri}(Z_1, R_1, Z_2)$ and $B_2 = \mathrm{Tri}(Z_1, R_2, Z_2)$ where R_1 and R_2 are positive. We will complete the proof by showing that $R_1 = R_2$.

Let $f_1(x) \neq 0$. Then $x \in M$ and the matrices $B_1(x)$ and $B_2(x)$ are unitarily equivalent. Also, $r_1(x)$, $r_2(x) \geqq 0$. By Lemma 8.1 $r_1(x) = r_2(x)$. Since this holds on a dense subset we have $r_1 = r_2$ and the theorem is proved.

In the course of this proof it has appeared that the reduced binormal operator (A_{ij}) which generates the ring $\boldsymbol{R}$ is unitarily equivalent in $\boldsymbol{R}$ to a matrix $\mathrm{Tri}(Z_1, R, Z_2)$ with positive definite R. This is, in general, far from being a canonical form since, as we have seen, Z_2 can be taken to be any solution of a fixed quadratic operator equation. It is clear, however, that with the choice of Z_2 once made the other entries in the matrix are uniquely determined.

It should also be noted explicitly that by adding one more term to the list of conditions of this theorem we can easily extend our result to include binormal operators which are not assumed to be reduced. In fact, if A_1 and A_2 are binormal operators with normal kernels N_1 and N_2 respectively, it

follows at once that A_1 and A_2 are unitarily equivalent if and only if there exist isomorphisms ϕ and ψ such that

$$1)\quad \phi^0(f_i(A_2)) = f_i(A_1) \quad (i = 1, \cdots, 4); \qquad 2)\quad \psi^0(N_2) = N_1.$$

(In general, of course, ϕ and ψ are defined on different Hilbert spaces.) In the special cases treated in § 7, where it was found that a single operator suffices to generate the center of $\boldsymbol{R}$, we may expect to find relations among these functions f_i, and indeed it is easy to verify that if $A^2 = 0$ we have $f_4(A) = f_2(A) = 0$ and $f_3(A) = f_1(A) = (A^*A + AA^*)^2$, while, on the other hand, if $A^2 = A$ the relations $f_4(A) = 0$ and $\frac{1}{4}f_3(A) = f_2(A) = f_1(A) = (A - A^*)^4 - (A - A^*)^2$ obtain. Thus the results of the special cases are implied in the general result but in a relatively complicated form, as could be expected from the fact that we have had to raise the degree of the polynomials f_i to take care of the most general 2×2 matrix.

10. Conclusion. Dixmier [2] has given a solution of the problem of the simultaneous unitary equivalence of two projections, or, what is the same thing, of the unitary equivalence problem for operators of the form $E + iF$ where E and F are projections. Now the operator $E + iF$ generates the same ring as do E and F together, and hence is binormal by Lemma 7. 1. We shall now use this fact to solve the problem anew, arriving at essentially the same result as that announced in [2].

Let E and F be projections on $\mathfrak{H}$. The first step is to reduce the problem to the case where the ring $\boldsymbol{R}$ generated by E and F is homogeneous, and since the non-normal part of $\boldsymbol{R}$ cannot here be identified in terms of a normal kernel, we must describe the subspace $\mathfrak{M}$ directly (cf. Theorem 1). The projection G of $\mathfrak{H}$ onto $\mathfrak{M}$ is, as we know, the sum of P, the maximal central abelian projection in $\boldsymbol{R}$, and the annihilating projection for $\boldsymbol{R}$, which is readily seen to be $(1 - E) \wedge (1 - F)$. The first problem before us then is to identify P.

To begin with, a projection H in the center $\boldsymbol{Z}$ of $\boldsymbol{R}$ is abelian in $\boldsymbol{R}$ if HE and HF commute. In particular then $E \wedge F$, $E \wedge (1 - F)$ and $(1 - E) \wedge F$ are abelian projections. Let us denote by Q the central abelian projection

$$(E \wedge F) \vee ((1 - E) \wedge F) \vee (E \wedge (1 - F)) = E \wedge F + (1 - E) \wedge F + E \wedge (1 - F).$$

Then of course $Q \leqq P$. But on the other hand

$$1 = EF + (1 - E)F + E(1 - F) + (1 - E)(1 - F)$$

and, making use of the properties ascribed to P, we see that

$$P = P(E \wedge F) + P((1-E) \wedge F) + P(E \wedge (1-F)) + P((1-E) \wedge (1-F))$$
$$= PQ$$

(note that $P \varepsilon \boldsymbol{R}$ implies $P((1-E) \wedge (1-F)) = 0$) whence $P \leqq Q$. Thus we have proved that $P = (E \wedge F) + (1-E) \wedge F + E \wedge (1-F)$, and hence also that

$$G = (E \wedge F) + (1-E) \wedge F + E \wedge (1-F) + (1-E) \wedge (1-F).$$

We assume accordingly that E and F are projections on $\mathfrak{H}$ for which

$$E \wedge F = (1-E) \wedge F = E \wedge (1-F) = (1-E) \wedge (1-F) = 0$$

and we proceed to investigate the circumstances under which E and F are simultaneously unitarily equivalent to another pair of projections, say E_1 and F_1, for which this requirement is also fulfilled. (The reader will perceive that this initial simplification of the problem is exactly the one made by Dixmier. In the terminology of [2] we are assuming that the subspaces onto which E and F project are "in position p" relative to one another.) The ring $\boldsymbol{R}$ is then a matrix ring which we represent on $M_2(\mathfrak{X})$ in the usual way. The functions $E(\cdot)$ and $F(\cdot)$ corresponding to the generating projections E and F are projection valued at every point. Let $t(x)$ denote the trace of $E(x)$. Then t can take only the values 0, 1, and 2, and the relations $t(x) = 0$ or $t(x) = 2$ correspond to the situations $E(x) = 0$ and $E(x) = 1$. Since the set of points for which $E(x)F(x) \neq F(x)E(x)$ must be dense we conclude that $t = 1$, and hence that $E(x)$ is singular everywhere. It is now a simple matter to construct a unitary function $U(\cdot)$ such that

$$U(x)E(x)U(x)^* = E'(x) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

for every $x \varepsilon \mathfrak{X}$. (We may in fact use the same function defined in the first example of § 7.) Then

$$U(x)F(x)U(x)^* = F'(x) = \begin{pmatrix} a(x) & c(x) \\ \bar{c}(x) & 1-a(x) \end{pmatrix}.$$

The function a is real and $c\bar{c} = a - a^2$. Let γ be a continuous function on $\mathfrak{X}$ such that $|c(x)| \cdot \gamma(x) = c(x)$. (That such a γ exists can be shown as usual by returning to the ring $\boldsymbol{Z}'$ which is isomorphic to $C(\mathfrak{X})$ and expresssing the operator C corresponding to c in polar form.) If $c(x) \neq 0$ we must have $|\gamma(x)| = 1$, and since c cannot vanish on any open set we conclude that γ

has absolute value 1 everywhere. Thus the function $U'(\cdot)$ defined by $U'(x) = \mathrm{Diag}(1, \gamma(x))$ is unitary in $M_2(\mathfrak{X})$, and writing

$$F''(x) = U'(x)F'(x)U'(x)^*$$

we see

$$F''(x) = \begin{pmatrix} a(x) & |\,c(x)\,| \\ |\,c(x)\,| & 1-a(x) \end{pmatrix}$$

while $E''(x) = U'(x)E'(x)U'(x)^* = E'(x)$. Since $|\,c(x)\,| = (a(x) - a(x)^2)^{\frac{1}{2}}$ one verifies that for every x

$$(E''(x) + F''(x) - 1)^2 = (E(x) + F(x) - 1)^2 = \begin{pmatrix} a(x) & 0 \\ 0 & a(x) \end{pmatrix}$$

from which it follows as usual that $(E + F - 1)^2$ generates the center $\boldsymbol{Z}$ of $\boldsymbol{R}$. By pursuing the same line of reasoning as we have used in each previous instance we arrive at the result: *If E_i and F_i are projections on Hilbert spaces $\mathfrak{H}_i$ for $i = 1, 2$ and if*

$$E_i \wedge F_i = E_i \wedge (1 - F_i) = (1 - E_i) \wedge F_i = (1 - E_i) \wedge (1 - F_i) = 0$$

then E_1 and F_1 are simultaneously unitarily equivalent to E_2 and F_2 (order is of no importance) if and only if the self-adjoint operators $(E_1 + F_1 - 1)^2$ and $(E_2 + F_2 - 1)^2$ are unitarily equivalent.

11. In connection with the treatment of binormal rings and operators given above certain questions naturally come to mind. Among these, most notably, are the following:

1) How, and to what extent, can this work be generalized to include rings which are $n \times n$ matrix rings for $n > 2$, and the operators which generate such rings? Can one go on at once to consider "trinormal" and, more generally, "n-normal" operators?

2) To what extent can the set of unitary invariants given in Theorem 5 be simplified?

With regard to 1), the fact is that there is nothing in § 2-§ 5 in which the number 2 plays any essential role. If a ring is subjected to any polynomial identity whatever, the above line of reasoning can be pursued with only unessential changes. The relevant facts are as follows: If $\boldsymbol{R}$ is a ring on $\mathfrak{H}$ satisfying a polynomial identity, then for some (finite) cardinal number n (depending on the polynomial identity) $\boldsymbol{R}$ has a direct factorization $\boldsymbol{R} = \boldsymbol{R}_1 \oplus \cdots \oplus \boldsymbol{R}_n$ corresponding to a decomposition $\mathfrak{H} = \mathfrak{H}_1 \oplus \cdots \oplus \mathfrak{H}_n$. The ring $\boldsymbol{R}_1$ on $\mathfrak{H}_1$ is the normal part separated out in Theorem 1 and, for

each $j > 1$, $\boldsymbol{R}_j$ on $\mathfrak{M}_j$ is "homogeneous of order j" in the sense that, in $\boldsymbol{R}_j$, 1 is the sum of j orthogonal, equivalent, abelian projections. For each homogeneous part $\boldsymbol{R}_j$ the analogs of Theorems 2, 3 and 4 hold, and the proofs need scarcely be modified. (In particular, Theorem 3 provides a set of unitary invariants for rings satisfying a polynomial (or finite rings of type I in the terminology of [8])). Also, n can be chosen at will by making an appropriate choice of polynomial identity, just as (*) was selected to make $n = 2$, and hence it is possible to give definitions of n-normal operators for any finite n. It would appear that for each n the solution of the unitary equivalence problem awaits only an adequate and applicable solution of the problem for the special case of $n \times n$ matrices. It should be observed at this point, however, that a straightforward generalization of the method we have used in the case $n = 2$ would involve setting forth a suitably separating collection of polynomials on the matrix ring which take their values in the center. As of this writing no such center valued polynomials are known to the author in any case $n > 2$.

If, on the other hand, we drop the assumption that $\boldsymbol{R}$ satisfy a polynomial identity, and turn rather to the direct assumption that there be an adequate supply of abelian projections (assuming, in other words, that $\boldsymbol{R}$ is type I but not necessarily finite) the case for the results of this paper disintegrates at once. The above mentioned analog of Theorem 1 holds still, but of course n need no longer be finite. (A ring is "homogeneous of order $\aleph$" if 1 is the sum (supremum) of $\aleph$ equivalent, orthogonal, abelian projections in it.) The very meanings of Theorems 2 and 4, depending as they do on matricial concepts, are in serious jeopardy, while Theorem 3, which is meaningful, is, unfortunately, false. Indeed, it is not hard to construct a non-trivial factor on $\aleph$-dimensional Hilbert space which is homogeneous of order $\aleph$ (provided, of course, that $\aleph$ is infinite).[7] These rings have been successfully investigated in [13], Part II, where a set of unitary invariants is given which does not depend on finiteness, and more recently in [6] where it is shown, among other things, that the algebraic analog of Theorem 3 remains true.

With regard to question 2), we shall content ourselves with showing by an example that in the presence of any non-trivial center it is not possible to do away with the simultaneity of the unitary equivalence of $f_i(A)$ and $f_i(B)$. Let t_i and d_i be any complex numbers for $i = 1, 2$ and let a_i and b_i be so chosen that $a_i + b_i = t_i$ and $a_i b_i = d_i$. Choose also numbers p_i and q_i for

[7] On the direct sum of $\aleph$ copies of a 2-dimensional space $\mathfrak{K}$ let $\boldsymbol{R}$ be the commutant of the $\aleph$-fold copy of the ring of all operators on $\mathfrak{K}$.

$i=1,2$ such that $p_1+q_1=t_1$ and $p_1q_1=d_2$ while $p_2+q_2=t_2$ and $p_2q_2=d_1$. Let m be any positive number exceeding the four quantities $|a_i-b_i|^2$ and $|p_i-q_i|^2$. Then there exist positive real numbers r_i and s_i such that, for $i=1,2$, $r_i^2(r_i^2+|a_i-b_i|^2)=s_i^2(s_i^2+|p_i-q_i|^2)=m$. Let $Z_i=\mathrm{Tri}(a_i, r_i, b_i)$ and $Y_i=\mathrm{Tri}(p_i, s_i, q_i)$ and define the two binormal 4×4 matrices A and B by

$$A=\begin{pmatrix} Z_1 & 0 \\ 0 & Z_2 \end{pmatrix}, \qquad B=\begin{pmatrix} Y_1 & 0 \\ 0 & Y_2 \end{pmatrix}.$$

If we make the further assumption that $t_1\neq t_2$ and $d_1\neq d_2$, we may assert that A and B are not unitarily equivalent, and similarly that Z_1 is not unitarily equivalent to Z_2, nor Y_1 to Y_2, from which it readily follows that A and B generate the same ring (all $\mathrm{Diag}(C, D)$ where C and D are any 2×2 matrices) whose center is represented on a 2-point space. Clearly neither A nor B has a normal kernel. But though they are not unitarily equivalent, it is an immediate consequence of the construction that $f_i(A)=f_i(B)$ for $i=1,2,3$ while the other functions, $f_4(A)=\mathrm{Diag}(d_1m, d_1m, d_2m, d_2m)$ and $f_4(B)=\mathrm{Diag}(d_2m, d_2m, d_1m, d_1m)$ are obviously unitarily equivalent.

Finally, a slight variation of this example can be used to show that, in proving that two reduced binormal operators which generate the same ring are unitarily equivalent in the ring, one cannot avoid making some restrictions (such as requiring that certain functions of them agree) which serve to exclude the intervention of an isomorphism corresponding to a homeomorphism of the representation space. For, in the above construction, with $t_1=t_2$ and $d_1\neq d_2$, it is readily seen that A and B are unitarily equivalent, but that they generate in common the same ring as before which obviously cannot contain U carrying A onto B.

RICE INSTITUTE.

REFERENCES.

[1] A. S. Amitsur and J. Levitzki, "Minimal identities for algebras," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 449-463.

[2] J. Dixmier, "Position relative de deux variétes linéaires fermées dans un espace de Hilbert," *Revue Scientifique*, vol. 86 (1948), pp. 387-399.

[3] ———, "Les anneaux d'opérateurs de classe finie," *Annales Scientifique de l'Ecole Normale Superieure*, vol. 66 (1949), pp. 209-261.

[4] P. R. Halmos, *Introduction to Hilbert space and the theory of spectral multiplicity*, New York, 1951.

[5] N. Jacobson, *The theory of rings, Mathematical Surveys*, 2. *American Mathematical Society*, New York, 1943.

[6] I. Kaplansky, "Algebras of type I," *Annals of Mathematics*, vol. 56 (1952), pp. 460-472.

[7] ———, "Normed algebras," *Duke Mathematical Journal*, vol. 16 (1949), pp. 399-418.

[8] ———, "Projections in Banach algebras," *Annals of Mathematics*, vol. 53 (1951), pp. 235-249.

[9] ———, "Symmetry of Banach algebras," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 396-399.

[10] F. J. Murray and J. v. Neumann, "On rings of operators," *Annals of Mathematics*, vol. 37 (1936), pp. 116-229.

[11] B. v. Sz.-Nagy, *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes, Ergebnisse der Mathematik*, 5, Berlin, 1942.

[12] J. v. Neumann, "Zur Algebra der Funktionaloperationen und Theorie der normalen Operatoren," *Mathematische Annalen*, vol. 102 (1930), pp. 370-427.

[13] I. E. Segal, *Decompositions of operator algebras, Memoirs*, 9, *American Mathematical Society*, New York, 1951.

[14] ———, "Equivalences of measure spaces," *American Journal of Mathematics*, vol. 73 (1951), pp. 275-313.

[15] M. H. Stone, *Linear transformations in Hilbert space and their applications to analysis, Collquium Publications*, 15, *American Mathematical Society*, New York, 1932.

[16] ———, "Boundedness properties in function lattices," *Canadian Journal of Mathematics*, vol. 1 (1949), pp. 176-186.

ON THE ALGEBRAS FORMED BY THE CAYLEY-DICKSON PROCESS.* [1]

By R. D. SCHAFER.

The iterative process by which L. E. Dickson obtained generalizations of Cayley's 8-dimensional non-associative algebra from the 4-dimensional algebras of generalized quaternions can be used, as A. A. Albert has shown in [1], to obtain central simple non-associative algebras $\mathfrak{A}_t$ of dimension 2^t $(t = 2, 3, 4, \cdots)$ over an arbitrary field F.

Let $\mathfrak{A}$ be an algebra of dimension n over F with identity element 1, and let an involution $x \to \bar{x}$ be defined in $\mathfrak{A}$ satisfying

$$(1) \qquad x + \bar{x} = t(x)1, \qquad x\bar{x} = \bar{x}x = n(x)1, \qquad t(x), n(x) \text{ in } F.$$

Then, for a fixed $\gamma \neq 0$ in F, an algebra $\mathfrak{B} = \mathfrak{A}\{\gamma\}$ of dimension $2n$ over F may be constructed as the set of all pairs $z = (x, y)$ for x, y in $\mathfrak{A}$, with the obvious addition and multiplication by scalars, and with multiplication defined by

$$(2) \qquad z_1 z_2 = (x_1, y_1)(x_2, y_2) = (x_1 x_2 + \gamma \bar{y}_2 y_1, y_2 x_1 + y_1 \bar{x}_2).$$

The pairs $(x, 0)$ may be identified with the elements of $\mathfrak{A}$, and the element $1 = (1, 0)$ is an identity element for $\mathfrak{B}$. The mapping $z = (x, y) \to \bar{z} = (\bar{x}, -y)$ is an involution of $\mathfrak{B}$ satisfying $z + \bar{z} = t(z)1$, $z\bar{z} = \bar{z}z = n(z)1$, where $t(z) = t(x)$, $n(z) = n(x) - \gamma n(y)$. Beginning with a two-dimensional commutative associative algebra $\mathfrak{A}_1$ which is either a separable quadratic extension of F or the semisimple algebra $F \oplus F$, and in which $x \to \bar{x}$ is the unique non-trivial automorphism, one obtains in this way a class of 2^t-dimensional algebras $\mathfrak{A} = \mathfrak{A}_t = \mathfrak{A}_1\{\gamma_2, \cdots, \gamma_t\}$ over F. We write $\mathfrak{B} = \mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma_{t+1}\} = \mathfrak{A}_1\{\gamma_2, \cdots, \gamma_{t+1}\}$, and it is clear that in our notation $\mathfrak{A}_i = \mathfrak{A}_1\{\gamma_2, \cdots, \gamma_i\}$ and $\mathfrak{A}_t = \mathfrak{A}_i\{\gamma_{i+1}, \cdots, \gamma_t\}$ imply $\mathfrak{A}_t = \mathfrak{A}_1\{\gamma_2, \cdots, \gamma\}$. The best-known cases of the algebras $\mathfrak{A}_t$ are the algebras $\mathfrak{Q} = \mathfrak{A}_2$ of generalized quaternions, and the Cayley-Dickson algebras $\mathfrak{C} = \mathfrak{A}_3$.

In this paper we derive first some elementary properties of the algebras $\mathfrak{A}_t$. Our chief purpose is to compute their derivation algebras $\mathfrak{D}(\mathfrak{A}_t)$ for

* Received June 2, 1953.

[1] This research was supported in part by a grant from the National Science Foundation.

$t > 3$. The derivation algebras $\mathfrak{D}(\mathfrak{A}_t)$ for $t = 2$ and 3 are well-known. $\mathfrak{D}(\mathfrak{Q})$ is isomorphic to the 3-dimensional Lie algebra $[\mathfrak{Q}, \mathfrak{Q}]$, while any $\mathfrak{L} = \mathfrak{D}(\mathfrak{C})$ of characteristic $\neq 2$ is 14-dimensional and, for F of characteristic 0, is a central simple Lie algebra of type G in the sense that $\mathfrak{L}_\Omega$ is the exceptional simple Lie algebra G_2 over Ω, the algebraic closure of F [3]. We show, for $t \geqq 3$ and characteristic $\neq 2, 3$, that $\mathfrak{D}(\mathfrak{A}_t\{\gamma\}) \cong \mathfrak{D}(\mathfrak{A}_t)$. Hence, for all $t > 3$, $\mathfrak{D}(\mathfrak{A}_t) \cong \mathfrak{D}(\mathfrak{A}_3)$ is 14-dimensional, a central simple Lie algebra of type G in case F is of characteristic 0.

Added in the proofs (*January 20, 1954*). I. Kaplansky announced at the American Mathematical Society's 1953 Summer Institute on Lie algebras and Lie groups that $\mathfrak{D}(\mathfrak{A}_3)$ is central simple for F of characteristic $\neq 3$ (and not simple for characteristic 3). If we say that these 14-dimensional central simple Lie algebras of characteristic $\neq 3$ are of type G, we can make the obvious changes in our statements above and in the statement of Theorem 4.

1. Elementary properties of the algebras $\mathfrak{A}_t$. It is well-known that the algebras $\mathfrak{Q} = \mathfrak{A}_2$ are associative, and that the algebras $\mathfrak{C} = \mathfrak{A}_3$ are alternative [2] but not associative. Albert has shown [1, p. 172] that $\mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$ is alternative if and only if $\mathfrak{A}_t$ is associative. Hence the algebras $\mathfrak{A}_t$ are not alternative for $t > 3$. However, we shall show that all of the algebras $\mathfrak{A}_t$ are *flexible*; that is,

$$(xy)x = x(yx) \qquad \text{for all } x, y \text{ in } \mathfrak{A}_t.$$

The right and left multiplications R_x and L_x of an algebra $\mathfrak{A}$ are the mappings $y \to yx = yR_x$ and $y \to xy = yL_x$ respectively. If $S\colon x \to \bar{x} = xS$ is an involution of $\mathfrak{A}$, then $S^2 = I$ and

$$R_xS = SL_{xS}, \qquad L_xS = SR_{xS} \qquad \text{for every } x \text{ in } \mathfrak{A}. \tag{3}$$

The condition for flexibility of an algebra $\mathfrak{A}$ is

$$L_xR_x = R_xL_x \qquad \text{for every } x \text{ in } \mathfrak{A}. \tag{4}$$

Linearized, this identity becomes

$$L_xR_y + L_yR_x = R_xL_y + R_yL_x \qquad \text{for all } x, y \text{ in } \mathfrak{A}. \tag{5}$$

[2] The identities $x^2y = x(xy)$, $yx^2 = (yx)x$ are satisfied.

By (2) matrices for the right and left multiplications of $\mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$ may be written in terms of those for $\mathfrak{A}_t$ as:

$$R_z = \begin{pmatrix} R_x & L_y \\ \gamma L_{y^S} & R_{xS} \end{pmatrix} \qquad L_z = \begin{pmatrix} L_x & SL_y \\ \gamma SR_y & R_x \end{pmatrix}. \tag{6}$$

LEMMA 1. *If an algebra $\mathfrak{A}_t$ is flexible, then*

$$R_{y^S}R_y = L_{y^S}L_y = L_yL_{y^S} = R_yR_{y^S} \qquad \text{for every } y \text{ in } \mathfrak{A}_t. \tag{7}$$

For, applying (5) to $\bar{y}$, we obtain $(x\bar{y})y + n(y)x = y(\bar{y}x) + n(y)x$, or $R_{y^S}R_y = L_{y^S}L_y$. But $L_{y^S}L_y = t(y)L_y - L_y^2 = L_yL_{y^S}$, and similarly $R_{y^S}R_y = R_yR_{y^S}$.

LEMMA 2. *If an algebra $\mathfrak{A}_t$ is flexible, then*

$$t(xy) = t(yx), \qquad t((xa)y) = t(x(ay)) \qquad \text{for all } x, a, y \text{ in } \mathfrak{A}_t. \tag{8}$$

The first of these equalities does not require flexibility. To prove the second, we use (5) to compute

$$\begin{aligned} t((xa)y) - t(x(ay)) &= (xa)y + \bar{y}(\bar{a}\bar{x}) - x(ay) - (\bar{y}\bar{a})\bar{x} \\ &= a([L_x, R_y] + S[R_{xS}, L_{y^S}]) = a(I + S)[L_x, R_y] \\ &= t(a)(xy - xy) = 0. \end{aligned}$$

THEOREM 1. *The algebras $\mathfrak{A}_t$ are flexible for every t.*

This is known for $t = 1, 2, 3$. Proceeding in a proof by induction on t, we show that $[R_z, L_z] = 0$ for every z in $\mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$ from (6) as follows. We verify that

$$[R_x, L_x] + \gamma(L_ySR_y - SL_yL_{y^S}) = 0, \tag{9}$$

$$R_xSL_y + L_yR_x - L_xL_y - SL_yR_{xS} = 0, \tag{10}$$

$$\gamma(L_{y^S}L_x + R_{xS}SR_y - SR_yR_x - R_xL_{y^S}) = 0, \tag{11}$$

$$\gamma(L_{y^S}SL_y - SR_yL_y) + [R_{xS}, R_x] = 0 \tag{12}$$

for all x, y in $\mathfrak{A}_t$. Now (9) follows from (4), (3), and (7). Also

$$L_yR_x - L_xL_y + S(L_{xS}L_y - L_yR_{xS}) = (I + S)(L_yR_x - L_xL_y) = 0$$

since, applied to any a in $\mathfrak{A}_t$, it gives $t(a)(yx - yx) = 0$, proving (10). Moreover,

$$L_{y^S}L_x - R_xL_{y^S} + (R_{xS}L_{y^S} - L_{y^S}L_{xS})S = (L_{y^S}L_x - R_xL_{y^S})(I + S) = 0$$

since, applied to any a in $\mathfrak{A}_t$, it gives $t(x(\bar{y}a) - \bar{y}(ax)) = 0$ by (8). This proves (11). Finally (12) follows from (7) and (3).

The *center* of a non-associative algebra $\mathfrak{A}$ is the set of all elements c in $\mathfrak{A}$ satisfying

$$(13) \qquad xc = cx \qquad \text{for every } x \text{ in } \mathfrak{A},$$

$$(14) \qquad (cx)y = c(xy),\ (xc)y = x(cy),\ (xy)c = x(yc) \qquad \text{for all } x, y \text{ in } \mathfrak{A}.$$

The elements $\alpha 1$ are elements of the center of any algebra $\mathfrak{A}_t$. Furthermore $F1$ is the center of $\mathfrak{A}_t$ for $t \geqq 2$. This is well-known for the quaternion algebras $\mathfrak{Q} = \mathfrak{A}_2$ and the Cayley-Dickson algebras $\mathfrak{C} = \mathfrak{A}_3$, and for $t > 3$ is a consequence of the following lemma.

LEMMA 3. *Let c be an element of an algebra $\mathfrak{A}_t$ for $t \geqq 3$. Any one of the following conditions is sufficient for c to be in $F1$:*

$$(13') \qquad R_c = L_c,$$

and, for every z in $\mathfrak{A}_t$,

$$(14a) \quad L_c R_z = R_z L_c, \qquad (14b) \quad R_c R_z = R_{cz}, \qquad (14c) \quad R_z R_c = R_{zc}.$$

For the Cayley-Dickson algebras $\mathfrak{A}_3$ the equivalence of (14a), (14b), (14c) follows from the alternative law. Also it is well-known in this case that either (13) or (14) suffices for c to be in $F1$. Assume that any one of the four conditions on an element of $\mathfrak{A}_t$ for $t \geqq 3$ guarantees that it is in $F1$. Take $c = (a, b)$ in $\mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$. Then (13′) and (6) imply in particular that $R_a = L_a$, $L_{bS} = SR_b$, $L_b = SL_b$. Then we have $a = \alpha 1$ by the assumption of the induction. Also $b = b$ implies $b = \beta 1$, so that $\beta I = \beta S$. Then $\beta = 0$, $c = (\alpha 1, 0)$ in $F1$. Next (14a) and (6) imply in particular that $L_a R_x = R_x L_a$, $SL_b L_{yS} = L_y S R_b$ for all x, y in $\mathfrak{A}_t$. Then $a = \alpha 1$ by the assumption of the induction. By (3) we have

$$(15) \qquad L_b L_{yS} = R_{yS} R_b \qquad \text{for every } y \text{ in } \mathfrak{A}_t.$$

Hence $L_b = R_b$ so that $b = \beta 1$. Then (15) becomes $\beta L_{yS} = \beta R_{yS}$ for every y in $\mathfrak{A}_t$. Since there exists y in $\mathfrak{A}_t$ such that $L_{yS} \neq R_{yS}$, we have $\beta = 0$, and $c = (\alpha 1, 0)$ in $F1$. Similarly (14b) or (14c) implies that c is in $F1$.

It follows from (1) that $x^2 - t(x)x + n(x)1 = 0$ for every x in $\mathfrak{A}_t$. We say that $\mathfrak{A}_t$ has degree two. The algebras $\mathfrak{A}_t$ are power-associative, since it is easy to see that every non-associative algebra of degree two is power-associative.

We assume throughout the remainder of the paper that the characteristic of F is not two. Then $\mathfrak{A}_1$ itself is formed by the Cayley-Dickson process from $\mathfrak{A}_0 = F$ where $\xi \to \bar{\xi} = \xi$ is the identity transformation on $\mathfrak{A}_0$. For we have $\mathfrak{A}_1 = \mathfrak{A}_0\{\gamma_1\}$, $\gamma_1 \neq 0$ in F.

A normalized basis for $\mathfrak{A}_t$ over F may be chosen as in the known cases $t = 1, 2, 3$. Let $m = 2^t$. Beginning with a basis $1, f_1$ for $\mathfrak{A}_1$ in which $f_1^2 = \alpha_1 1$, $\alpha_1 = \gamma_1 \neq 0$, it follows readily by induction that there is a basis $1, f_1, f_2, \cdots, f_{m-1}$ for $\mathfrak{A}_t$ satisfying

$$f_i^2 = \alpha_i 1, \qquad \alpha_i \neq 0 \text{ in } F, \ i = 1, \cdots, m-1, \tag{16}$$

and, for $i \neq j$,

$$f_i f_j = -f_j f_i = \beta_{ij} f_k, \qquad \beta_{ij} \neq 0 \text{ in } F, \ i, j = 1, \cdots, m-1, \tag{17}$$

the elements α_i and β_{ij} depending on $\gamma_1, \gamma_2, \cdots, \gamma_t$, and f_k being an element $(\neq 1, f_i, f_j)$ of the basis uniquely determined by f_i and f_j. If $x = \xi_0 1 + \sum \xi_i f_i$, then $\bar{x} = \xi_0 1 - \sum \xi_i f_i$, $t(x) = 2\xi_0$, $n(x) = \xi_0^2 - \sum \alpha_i \xi_i^2$. In our notation, if $\mathfrak{A} = \mathfrak{A}_t$, $\mathfrak{B} = \mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$, and $r = 2^{t+1} = 2m$, then for $i = m, m+1, \cdots, r-1$, we have $f_i = (0, f_{i-m})$ if we make the convention that $f_0 = 1$, and we also have

$$\alpha_m = \gamma, \qquad \alpha_i = -\gamma \alpha_{i-m} \qquad \text{for } i = m+1, \cdots, r-1. \tag{18}$$

It follows from (16) and (17) that the algebras $\mathfrak{A}_t$ are Jordan-admissible, for the algebra $\mathfrak{A}_t^+$ in which multiplication is defined by $x \cdot y = \frac{1}{2}(xy + yx)$ is a special central simple Jordan algebra of degree two. Thus the algebras $\mathfrak{A}_t$ are flexible central simple Jordan-admissible algebras of degree two.[3]

The following lemmas will be used in the next section to prove Theorem 3. Although $\mathfrak{A}_t$ is not alternative for $t > 3$, the alternative law is satisfied by all pairs of elements in the normalized basis: $f_i^2 f_j = f_i(f_i f_j)$, $f_j f_i^2 = (f_j f_i) f_i$, as may be seen from the following lemma.

LEMMA 4. *In any algebra $\mathfrak{A}_t$ with basis $1, f_1, f_2, \cdots, f_{m-1}$ satisfying (16) and (17), we have*

$$f_i(f_i x) = \alpha_i x = (x f_i) f_i \qquad \text{for every } x \text{ in } \mathfrak{A}_t. \tag{19}$$

It is sufficient to prove

$$f_i(f_i f_j) = \alpha_i f_j = (f_j f_i) f_i \qquad \text{for } i \neq j. \tag{20}$$

Now (17) implies $(f_i f_j) f_i + (f_j f_i) f_i = 0$ and $f_i(f_i f_j) + f_i(f_j f_i) = 0$. Subtracting, and using Theorem 1, we have $f_i(f_i f_j) = (f_j f_i) f_i$. We prove

[3] In [2, p. 588] Albert studies flexible central simple Jordan-admissible algebras of degree two. We note that this class includes the non-commutative algebras $\mathfrak{M}$ such that $\mathfrak{M}_K \cong \mathfrak{A}_t(\lambda)_K$ for some extension K of F, where $\mathfrak{A}_t(\lambda)$ is the algebra which is the same vector space as $\mathfrak{A}_t$ for $t \geqq 2$ and in which multiplication is defined by $x * y = \lambda xy + (1-\lambda) yx$ for $\lambda \neq 1/2$.

$f_i(f_if_j) = \alpha_i f_j$ by induction on t, assuming (20) for $\mathfrak{A}_t$. Then in $\mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$ there are three cases left to consider: (i) $1 \leqq i \leqq m-1$, $m \leqq j \leqq r-1$; (ii) $m \leqq i \leqq r-1$, $1 \leqq j \leqq m-1$; (iii) $m \leqq i \leqq r-1$, $m \leqq j \leqq r-1$. In case (i), we have $f_i(f_if_j) = (0, (f_{j-m}f_i)f_i) = (0, \alpha_i f_{j-m}) = \alpha_i f_j$. In case (ii), $f_i(f_if_j) = (\gamma(f_j\bar{f}_{i-m})f_{i-m}, 0) = \alpha_i f_j$ by (18), and case (iii) holds similarly.

LEMMA 5. *Let $f_ix = xf_i$ for some fixed i, $1 \leqq i \leqq m-1$. Then $x = \xi_0 1 + \xi_i f_i$.*

For $x = \xi_0 1 + \sum_j \xi_j f_j$ implies

$$[f_i, x] = \sum_{j \neq i} \xi_j [f_i, f_j] = 2 \sum_{j \neq i} \xi_j f_i f_j = 2f_i\Big(\sum_{j \neq i} \xi_j f_j\Big) = 0$$

by (17). Then $f_i^2 = \alpha_i 1 \neq 0$ implies $\sum_{j \neq i} \xi_j f_j = 0$, or $x = \xi_0 1 + \xi_i f_i$.

For any algebra $\mathfrak{A}_t$, $\frac{1}{2}t(x\bar{y})$ is a nondegenerate symmetric bilinear form $g(x, y)$ and we may define, for any linear transformation T on $\mathfrak{A}_t$, its *transpose* T' (relative to g) by $g(xT, y) = g(x, yT')$ for all x, y in $\mathfrak{A}_t$. For the unique algebra $\mathfrak{E}_t = \mathfrak{A}_0\{-1, \cdots, -1\}$ in which $\gamma_1 = \gamma_2 = \cdots = \gamma_t = -1$, we have $g(x, y) = \sum \xi_i \eta_i$ so that the matrix of T' relative to the normalized basis of $\mathfrak{E}_t$ is the transpose of the matrix of T. Then T is symmetric (skew-symmetric) if and only if its matrix is symmetric (skew-symmetric).

LEMMA 6. *For any $\mathfrak{A}_t$ we have*

$$R_z' = R_{z\bar{s}}, \qquad L_z' = L_{z\bar{s}}. \tag{21}$$

For,

$$g(x, yR_z') = g(xR_z, y) = \tfrac{1}{2}t((xz)\bar{y}) = \tfrac{1}{2}t(x(z\bar{y})) = g(x, y\bar{z}) = g(x, yR_{z\bar{s}})$$

implies $yR_z' = yR_{z\bar{s}}$, or $R_z' = R_{z\bar{s}}$. Similarly $L_z' = L_{z\bar{s}}$.

Finally we remark that in $\mathfrak{A}_t$ $(t \geqq 2)$ an element is a sum of commutators if and only if its trace is zero, so that any element x of $\mathfrak{A}_t$ $(t \geqq 2)$ may be written in the form

$$x = \xi_0 1 + \sum_i [a_i, b_i], \qquad x, a_i, b_i \text{ in } \mathfrak{A}_t, t \geqq 2. \tag{22}$$

It is sufficient to show that any basal element f_j is a commutator. For $i \neq j$ we have $f_if_j = \beta_{ij}f_k$, which implies $\alpha_i f_j = \beta_{ij}f_if_k = -\beta_{ij}f_kf_i$, by (19), since $k \neq i$. Hence $2\alpha_i f_j = [\beta_{ij}f_i, f_k]$.

2. Derivations of the algebras $\mathfrak{A}_t$. A *derivation* D of a non-associative algebra $\mathfrak{A}$ is a linear transformation on $\mathfrak{A}$ satisfying

$$(xy)D = (xD)y + x(yD) \qquad \text{for all } x, y \text{ in } \mathfrak{A}. \tag{23}$$

Equation (23) is equivalent to either one of the following:

$$[R_x, D] = R_{xD}, \quad [L_x, D] = L_{xD} \qquad \text{for every } x \text{ in } \mathfrak{A}. \tag{24}$$

The set of all derivations of $\mathfrak{A}$ is a Lie algebra $\mathfrak{D}(\mathfrak{A})$, called the *derivation algebra* of $\mathfrak{A}$.

LEMMA 7. *Let D be a derivation of an algebra $\mathfrak{A}_t$. Then*

$$t(xD) = 0 \text{ for every } x \text{ in } \mathfrak{A}_t, \tag{25}$$

$$SD = DS \tag{26}$$

where S is the involution $x \to \bar{x}$ of $\mathfrak{A}_t$, and

$$D' = -D. \tag{27}$$

Since $\mathfrak{A}_1$ is separable and commutative, it is known that in this case $D = 0$, so we may assume that $t \geqq 2$. From (22), since it is known that $1D = 0$, we have $xD = \sum_i [a_i, b_i]D = \sum_i ([a_iD, b_i] + [a_i, b_iD])$ so that $t(xD) = 0$. Also

$$xD + xDS = t(xD)1 = 0 = t(x)1D = (x + \bar{x})D = xD + xSD$$

for every x in $\mathfrak{A}_t$, implying (26). Finally (25) and (26) imply that $t((x\bar{y})D) = t((xD)\bar{y}) + t(x(yDS)) = 0$, or $g(xD, y) = -g(x, yD)$. Then $g(x, yD') = g(xD, y) = -g(x, yD)$, implying (27).

THEOREM 2. *Any derivation A of an algebra $\mathfrak{A}_t$ can be extended to a derivation D of $\mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$ by defining*

$$zD = (x, y)D = (xA, yA). \tag{28}$$

In fact, $[R_z, D]$ is

$$\left[\begin{pmatrix} R_x & L_y \\ \gamma L_{yS} & R_{xS} \end{pmatrix}, \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}\right] = \begin{pmatrix} [R_x, A] & [L_y, A] \\ \gamma[L_{yS}, A] & [R_{xS}, A] \end{pmatrix} = \begin{pmatrix} R_{xA} & L_{yA} \\ \gamma L_{ySA} & R_{xSA} \end{pmatrix}$$

$= R_{zD}$, by (24) and (26).

The following theorem is not true for $t = 1, 2$.

THEOREM 3. *Let F be of characteristic $\neq 2, 3$. Then, if $t \geqq 3$, every derivation D of $\mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$ has the form (28) for a derivation A of $\mathfrak{A}_t$.*

We begin by proving the result for $\mathfrak{E}_t = \mathfrak{A}_0\{-1, \cdots, -1\}$ and $\mathfrak{E}_{t+1} = \mathfrak{E}_t\{-1\}$. In this case (27) implies that the matrix of D is skew-symmetric:

$$D = \begin{pmatrix} A & B \\ -B' & C \end{pmatrix},$$

where $A' = -A$, $C' = -C$. Since $[R_z, D] = R_{zD}$ for every $z = (x, y)$ in $\mathfrak{E}_{t+1}$ we have in particular that, for every x and y in $\mathfrak{E}_t$,

(29) $[R_x, A] = R_{xA}$, (30) $-L_yB' + BL_{yS} = -R_{yB'}$,

(31) $-L_{yS}B + B'L_y = -R_{yB'S}$, (32) $[R_{xS}, C] = R_{xAS}$.

Equation (29) states that A is a derivation of $\mathfrak{E}_t$. Then (32) and (26) imply

(33) $[R_x, C] = R_{xA}$ for every x in $\mathfrak{E}_t$.

Then $xC - (1C)x = xA$, or

(34) $C = A + L_c, \quad c = 1C.$

Putting (34) in (33), we have $[R_x, A + L_c] = R_{xA} = R_{xA} + [R_x, L_c]$ by (24), or

$$[R_x, L_c] = 0 \qquad \text{for every } x \text{ in } \mathfrak{E}_t.$$

Since $t \geqq 3$, Lemma 3(14a) implies that c is in $F1$. But $C' = -C$ implies $g(1, 1C') = g(1C, 1) = g(c, 1) = c = -g(1, 1C) = -c$, or $c = 0$. Hence

(35) $C = A.$

Since $1D = 0 = (1A, 1B)$, we have

(36) $1B = 0.$

Put $y = 1$ in (30). Then

(37) $B' = B + R_c, \qquad c = 1B'.$

Transposition in (30) yields $-BL_{yS} + L_yB' = -R_{yB'S} = R_{yB'}$ by (21). Hence

(38) $B'S = -B'.$

Putting (37) into (30) and (31), we then have $[R_c, L_y] + BL_{y+yS} = L_{y+yS}B$, or $[R_c, L_y] = 0$ for every y in $\mathfrak{E}_t$. That is, $R_{xc} = R_xR_c$ for every x in $\mathfrak{E}_t$, and by Lemma 3(14c), $c = 1B'$ is in $F1$. Then (38) implies $\bar{c} = -c$, or $c = 0$, and B is symmetric by (37). Hence (30) becomes

(39) $L_yB = BL_{yS} + R_{yB}$ for every y in $\mathfrak{E}_t$.

Let $u = f_if_j$. Then (39) implies that $uB = -f_i(f_jB) + f_j(f_iB)$ and

$$\begin{aligned}(f_iu)B &= -f_i(uB) + u(f_iB)\\ &= -f_jB - f_i\{f_j(f_iB)\} + (f_if_j)(f_iB) = -f_jB\end{aligned}$$

by (19), or

$$(f_if_j)(f_iB) = f_i\{f_j(f_iB)\}, \qquad i, j = 1, \cdots, m-1. \tag{40}$$

Write for the $(i+1)$-st row vector of B

$$f_iB = (a_i, b_i) \qquad \text{for } a_i, b_i \text{ in } \mathfrak{E}_{t-1}. \tag{41}$$

Since B is symmetric, it follows from (36) that $t(a_i) = 0$. Let $h = 2^{t-1} = m/2$. For $i, j = 1, \cdots, h-1$, we have $(f_if_j)(f_iB) = ((f_if_j)a_i, b(f_if_j))$ and $f_i\{f_j(f_iB)\} = (f_i(f_ja_i), (b_if_j)f_i)$. Hence (40) implies in particular that

$$b_i(f_if_j) = (b_if_j)f_i, \qquad i, j = 1, \cdots, h-1. \tag{42}$$

Let $j = h$; $i = 1, \cdots, h-1$ in (40). Then $(-\bar{b}_if_i, f_i\bar{a}_i) = (-f_i\bar{b}_i, \bar{a}_if_i)$, so that

$$a_if_i = f_ia_i, \qquad f_ib_i = b_if_i, \qquad i = 1, \cdots, h-1.$$

By Lemma 5, $a_i = \lambda_i 1 + \mu_i f_i$, $b_i = \nu_i 1 + \pi_i f_i$ for $i = 1, \cdots, h-1$. But $t(a_i) = 0$ implies

$$a_i = \mu_i f_i, \qquad i = 1, \cdots, h-1. \tag{43}$$

For any $j \neq i$, (42) implies $\nu_i f_i f_j + \pi_i f_i(f_if_j) = \nu_i f_jf_i + \pi_i(f_if_j)f_i$, or $2(\nu_i f_i - \pi_i 1)f_j = 0$. Hence $\nu_i f_i - \pi_i 1 = 0$, $\nu_i = \pi_i = 0$,

$$b_i = 0, \qquad i = 1, \cdots, h-1. \tag{44}$$

Since B is symmetric, (36) and (44) imply

$$a_i = 0, \qquad i = h, \cdots, m-1. \tag{45}$$

For $i = h, \cdots, m-1$; $j = 1, \cdots, h-1$, equations (40), (41), (45) imply $(-\bar{b}_i(f_{i-h}\bar{f}_j), 0) = (-(\bar{f}_j\bar{b}_i)f_{i-h}, 0)$ so that

$$f_jb_h = b_hf_j, \qquad j = 1, \cdots, h-1, \tag{46}$$

and

$$(f_jf_{i-h})b_i = f_{i-h}(b_if_j), \qquad i = h+1, \cdots, m-1; \quad j = 1, \cdots, h-1. \tag{47}$$

Formula (46) implies that $xb_h = b_hx$ for every x in $\mathfrak{E}_{t-1}$ $(t-1 \geqq 2)$. By the definition of the center of the quaternion algebra $\mathfrak{E}_{t-1} = \mathfrak{E}_2$ if $t-1 = 2$, and by Lemma 3(13′) if $t-1 \geqq 3$, we have

$$b_h = \mu_h 1. \tag{48}$$

For $j = h$; $i = h+1, \cdots, m-1$, equations (40), (41), (45) imply $(0, -b_i f_{i-h}) = (0, -f_{i-h} b_i)$, or $b_i f_{i-h} = f_{i-h} b_i$, $i = h+1, \cdots, m-1$. Then it follows from Lemma 5 that $b_i = \lambda_i 1 + \mu_i f_{i-h}$ for $i = h+1, \cdots, m-1$. Put this in (47): $\lambda_i f_j f_{i-h} - \mu_i f_j = \lambda_i f_{i-h} f_j - \mu_i f_j$. Taking $j \neq i-h$, we have $\lambda_i = 0$ so that

$$b_i = \mu_i f_{i-h}, \qquad i = h+1, \cdots, m-1. \tag{49}$$

Combining (41), (43), (44), (45), (48), (49), we have

$$f_i B = \mu_i f_i, \qquad i = 1, \cdots, m-1, \tag{50}$$

or B is the diagonal matrix $B = \operatorname{diag}\{0, \mu_1, \cdots, \mu_{m-1}\}$.

Using (50), we see that (39) implies $(f_i f_j)B = -\mu_j f_i f_j + \mu_i f_j f_i$, or $\mu_k \beta_{ij} f_k = -\mu_j \beta_{ij} f_k - \mu_i \beta_{ij} f_k$. Hence

$$\mu_i + \mu_j + \mu_k = 0 \tag{51}$$

for any triple (i, j, k) satisfying $f_i f_j = \beta_{ij} f_k$. We shall prove, by induction on t ($t \geqq 3$), that the equations (51) imply $\mu_i = 0$ for $i = 1, \cdots, m-1$. For the Cayley algebra $\mathfrak{C} = \mathfrak{E}_3$ there are, without duplications, exactly seven triples (i, j, k). These are given by the seven quaternion subalgebras of $\mathfrak{C}$ with normalized bases which are contained in the basis $1, f_1, \cdots, f_7$ of $\mathfrak{C}$ [3, p. 775]. It is easily verified that the seven triples (i, j, k) are [4]

$$(1, 2, 3),\ (1, 4, 5),\ (1, 6, 7),\ (2, 4, 6),\ (2, 5, 7),\ (3, 4, 7),\ (3, 5, 6). \tag{52}$$

If F is of characteristic $\neq 2, 3$, the seven homogeneous equations (51) for (i, j, k) in (52) have only the trivial solution $\mu_1 = \mu_2 = \cdots = \mu_7 = 0$. Assume that in $\mathfrak{E}_{t-1}$ ($t > 3$) the only solution of the equations (51) is the trivial one

$$\mu_1 = \mu_2 = \cdots = \mu_{h-1} = 0. \tag{53}$$

Since the basis $1, f_1, \cdots, f_{m-1}$ for $\mathfrak{E}_t$ includes the basis $1, f_1, \cdots, f_{h-1}$ for $\mathfrak{E}_{t-1}$, we have (53) in $\mathfrak{E}_t$. Then $f_3 f_{h+1} = -f_{h+2}$, $f_i f_h = f_{h+i}$ ($i = 1, \cdots, h-1$) give h equations (51) for the triples

$$(3, h+1, h+2), \qquad (i, h, h+1), \qquad i = 1, \cdots, h-1. \tag{54}$$

By (53) the equations (51) for the triples (54) become $\mu_{h+1} + \mu_{h+2} = 0$, $\mu_h + \mu_{h+i} = 0$ ($i = 1, \cdots, h-1$), which clearly have only the trivial solution $\mu_h = \mu_{h+1} = \cdots = \mu_{m-1} = 0$ for F of characteristic $\neq 2$. Hence

[4] Our notation differs slightly from [3], but the desired triples (i, j, k) are the same.

$B = 0$, and by (35) any derivation D of $\mathfrak{E}_{t+1} = \mathfrak{E}_t\{-1\}$ has the form (28) where A is a derivation of $\mathfrak{E}_t$.

We extend this result to arbitrary $\mathfrak{A}_t$ and $\mathfrak{A}_{t+1} = \mathfrak{A}_t\{\gamma\}$ as follows. For A in $\mathfrak{D}(\mathfrak{E}_t)$, $A \leftrightarrow D$ in (28) is an isomorphism between the derivation algebras $\mathfrak{D}(\mathfrak{E}_t)$ and $\mathfrak{D}(\mathfrak{E}_{t+1})$. There is an extension K of F such that $\mathfrak{A}_{tK} \cong \mathfrak{E}_{tK}$, $\mathfrak{A}_{t+1\,K} \cong \mathfrak{E}_{t+1\,K}$. Hence

$$\mathfrak{D}(\mathfrak{A}_t)_K = \mathfrak{D}(\mathfrak{A}_{tK}) \cong \mathfrak{D}(\mathfrak{E}_{tK}) = \mathfrak{D}(\mathfrak{E}_t)_K \cong \mathfrak{D}(\mathfrak{E}_{t+1})_K \cong \mathfrak{D}(\mathfrak{A}_{t+1})_K.$$

For A in $\mathfrak{D}(\mathfrak{A}_t)$, $A \leftrightarrow D$ in (28) is an isomorphism of $\mathfrak{D}(\mathfrak{A}_t)$ onto a subalgebra $\mathfrak{D}$ of $\mathfrak{D}(\mathfrak{A}_{t+1})$. But $\mathfrak{D}(\mathfrak{A}_t)_K \cong \mathfrak{D}(\mathfrak{A}_{t+1})_K$ implies $\dim \mathfrak{D} = \dim \mathfrak{D}(\mathfrak{A}_t) = \dim \mathfrak{D}(\mathfrak{A}_{t+1})$, so that $\mathfrak{D} = \mathfrak{D}(\mathfrak{A}_{t+1})$, and every derivation D of $\mathfrak{A}_{t+1}$ has the form (28). This completes the proof of the theorem.

THEOREM 4. *Let F be of characteristic $\neq 2, 3$. Let $\mathfrak{C} = \mathfrak{A}_0\{\gamma_1, \gamma_2, \gamma_3\}$ be a Cayley-Dickson algebra over F and*

$$\mathfrak{A}_t = \mathfrak{C}\{\gamma_4, \cdots, \gamma_t\} = \mathfrak{A}_0\{\gamma_1, \gamma_2, \gamma_3, \cdots, \gamma_t\}$$

be any algebra formed from $\mathfrak{C}$ by iteration of the Cayley-Dickson process. Then $\mathfrak{D}(\mathfrak{A}_t) \cong \mathfrak{D}(\mathfrak{C})$, a 14-dimensional Lie algebra over F. Also $\mathfrak{C}$ is invariant under $\mathfrak{D}(\mathfrak{A}_t)$, and every derivation of $\mathfrak{C}$ can be extended in a unique way to a derivation of $\mathfrak{A}_t = \mathfrak{C}\{\gamma_4, \cdots, \gamma_t\}$. If F is of characteristic 0, then $\mathfrak{D}(\mathfrak{A}_t) \cong \mathfrak{D}(\mathfrak{C})$ is a central simple Lie algebra of type G.

It follows from Theorem 3 that $\mathfrak{D}(\mathfrak{A}_i\{\gamma_{i+1}\}) \cong \mathfrak{D}(\mathfrak{A}_i)$ for $i \geqq 3$. Then $\mathfrak{D}(\mathfrak{A}_t) \cong \mathfrak{D}(\mathfrak{A}_3) = \mathfrak{D}(\mathfrak{C})$. Also it follows from Theorem 3 that $\mathfrak{A}_i$ is invariant under $\mathfrak{D}(\mathfrak{A}_i\{\gamma_{i+1}\})$ for $i \geqq 3$. Hence $\mathfrak{C} = \mathfrak{A}_3$ is invariant under $\mathfrak{D}(\mathfrak{A}_t)$. That $\mathfrak{D}(\mathfrak{C})$ is 14-dimensional and, in case F has characteristic 0, is a central simple Lie algebra of type G is established in [3].

THEOREM 5. *Let F be of characteristic 0. Then the derivation algebras $\mathfrak{D}(\mathfrak{A}_0\{\gamma_1, \cdots, \gamma_t\})$ and $\mathfrak{D}(\mathfrak{A}_0\{\gamma_1', \cdots, \gamma_s'\})$ for $t \geqq 3$, $s \geqq 3$, are isomorphic if and only if the Cayley-Dickson subalgebras $\mathfrak{C} = \mathfrak{A}_0\{\gamma_1, \gamma_2, \gamma_3\}$ and $\mathfrak{C}' = \mathfrak{A}_0\{\gamma_1', \gamma_2', \gamma_3'\}$ are isomorphic.*

For $\mathfrak{D}(\mathfrak{A}_0\{\gamma_1, \cdots, \gamma_t\}) \cong \mathfrak{D}(\mathfrak{C})$ and $\mathfrak{D}(\mathfrak{A}_0\{\gamma_1', \cdots, \gamma_s'\}) \cong \mathfrak{D}(\mathfrak{C}')$ by Theorem 4. But $\mathfrak{D}(\mathfrak{C}) \cong \mathfrak{D}(\mathfrak{C}')$ if and only if $\mathfrak{C} \cong \mathfrak{C}'$ by [3, p. 782].

It is well-known that in the construction of a Cayley-Dickson algebra $\mathfrak{C}$ any generalized quaternion subalgebra $\mathfrak{Q}$ of $\mathfrak{C}$ will suffice for the algebra $\mathfrak{A}_2$ from which $\mathfrak{C} = \mathfrak{A}_3$ is constructed by the Cayley-Dickson process. That is, $\mathfrak{C} \cong \mathfrak{Q}\{\gamma\}$ for some γ in F [3, p. 776]. That this property is not enjoyed in general by the algebras $\mathfrak{A}_t$ for $t > 3$ may be seen from Theorem 5 (for F

of characteristic 0). For, let $\mathfrak{C} = \mathfrak{A}_0\{\gamma_1, \gamma_2, \gamma_3\}$ and $\mathfrak{C}' = \mathfrak{A}_0\{\gamma_1, \gamma_2, \gamma_4\}$ where $\mathfrak{C}'$ is not isomorphic to $\mathfrak{C}$ (as would be the case if F were the field of all real numbers and $\gamma_1 = \gamma_2 = \gamma_3 = -1$, $\gamma_4 = 1$). Then the algebra $\mathfrak{A}_4 = \mathfrak{A}_0\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\} = \mathfrak{C}\{\gamma_4\}$ contains a Cayley-Dickson subalgebra $\mathfrak{C}''$ isomorphic to $\mathfrak{C}'$. However, there is no γ in F such that $\mathfrak{A}_4 \cong \mathfrak{C}''\{\gamma\}$. For, if there were, then $\mathfrak{D}(\mathfrak{C}\{\gamma_4\}) \cong \mathfrak{D}(\mathfrak{C}''\{\gamma\})$ would imply $\mathfrak{C} \cong \mathfrak{C}'' \cong \mathfrak{C}'$ by Theorem 5. Hence $\mathfrak{A}_4$ cannot be constructed by the Cayley-Dickson process from its subalgebra $\mathfrak{C}''$. A similar argument holds for $t > 4$. It is likely that one can prove by a different method that the same situation exists for characteristic $p > 0$.

UNIVERSITY OF CONNECTICUT.

REFERENCES.

[1] A. A. Albert, "Quadratic forms permitting composition," *Annals of Mathematics*, vol. 43 (1942), pp. 161-177.

[2] ———, "Power-associative rings," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 552-593.

[3] N. Jacobson, "Cayley numbers and normal simple Lie algebras of type G," *Duke Mathematical Journal*, vol. 5 (1939), pp. 775-783.

A KRONECKER FACTORIZATION THEOREM FOR CAYLEY ALGEBRAS AND THE EXCEPTIONAL SIMPLE JORDAN ALGEBRA.*

By N. JACOBSON.[1]

A well known theorem, due to Wedderburn, states that if $\mathfrak{B}$ is an arbitrary associative algebra with an identity 1 and $\mathfrak{A}$ is a finite dimensional central simple subalgebra containing 1, then $\mathfrak{B}$ is the Kronecker product $\mathfrak{B} = \mathfrak{A} \otimes \mathfrak{U}$ where $\mathfrak{U}$ is the subalgebra of elements which commute with every $a \in \mathfrak{A}$. Recently Kaplansky indicated that the same result holds for $\mathfrak{B}$ any alternative algebra and $\mathfrak{A}$ an algebra of Cayley numbers.[2] It is the purpose of the present note to give a new proof of this result and to show that an analogous result holds for $\mathfrak{B}$ any Jordan algebra and $\mathfrak{A}$ an exceptional simple Jordan algebra of finite dimensions.

In considering a subalgebra $\mathfrak{A}$ of an algebra $\mathfrak{B}$ it is natural to regard $\mathfrak{B}$ as a bimodule relative to $\mathfrak{A}$. One is therefore led to believe that the underlying reason for the validity of the results of the type noted (beginning with Wedderburn's) is to be found in the theory of bimodules. This is indeed the case in the associative and alternative cases where the key result in the proof of the factorization theorem is the theorem that every associative or alternative bimodule for $\mathfrak{A}$ is completely reducible and $\mathfrak{A}$ has only one irreducible bimodule (cf. [2], § 18). We have recently proved that the analogous result holds also for any exceptional simple Jordan algebra of finite dimensions (cf. [2], § 18). However, there appear to be certain technical difficulties which have prevented our giving a similar type of proof for the Jordan theorem. Instead, we shall base our proof for these algebras on the corresponding result for Cayley algebras and on some structure theorems for Jordan algebras which were used in deriving the results on bimodules for Jordan algebras.

* Received November 9, 1953.

[1] The results of this paper were obtained while the author held a Guggenheim Memorial Fellowship.

[2] This is stated in [5]. Kaplansky observes that the proof can be obtained by using the arguments given by Albert in [1]. The situation for the other central simple alternative algebras is mentioned in the Concluding Remarks at the end of this paper.

1. The alternative theorem. Throughout this paper we assume that the base field is of characteristic $\neq 2$.

LEMMA 1. *Let $\mathfrak{B}$ be an arbitrary alternative algebra and $\mathfrak{A}$ a finite dimensional separable*[3] *subalgebra. Set $\mathfrak{U} = \{u \mid u \in \mathfrak{B}, au = ua, a \in \mathfrak{A}\}$. Then $\mathfrak{U}$ is a subalgebra of $\mathfrak{B}$ and $(au)(bv) = (ab)(uv)$ holds for all $a, b \in \mathfrak{A}$, $u, v \in \mathfrak{U}$.*

Proof. It suffices to prove this result under the added assumption that the base field is algebraically closed. In this case the structure theory shows that $\mathfrak{A}$ has a basis consisting of idempotent elements. Now let e be an idempotent element in $\mathfrak{B}$ and consider the Peirce decomposition

$$\mathfrak{B} = \mathfrak{B}_{11} + \mathfrak{B}_{10} + \mathfrak{B}_{01} + \mathfrak{B}_{00} \text{ where } \mathfrak{B}_{ij} = \{b_{ij} \mid eb_{ij} = \delta_{ij}b_{ij} = b_{ij}e\}.$$

If $u = \sum u_{ij}$, $u_{ij} \in \mathfrak{B}_{ij}$ then $eu = ue$ is equivalent to the condition that $u = u_{11} + u_{00}$. It is known that the subset $\mathfrak{U}(e)$ of elements of this form is a subalgebra. Since $\mathfrak{U} = \cap\mathfrak{U}(e)$, $e \in \mathfrak{A}$, $\mathfrak{U}$ is a subalgebra of $\mathfrak{B}$. Now let $x = \sum x_{ij}$, $x_{ij} \in \mathfrak{B}_{ij}$, be any element of $\mathfrak{B}$. If $u = u_{11} + u_{00}$, then

$$(xe)u = (x_{11} + x_{01})(u_{11} + u_{00}) = (x_{11} + x_{01})u_{11}$$

$$x(eu) = (x_{11} + x_{10} + x_{01} + x_{00})u_{11} = (x_{11} + x_{01})u_{11}.$$

Hence the associator $A(x, e, u) \equiv (xe)u - x(eu) = 0$ for all x in $\mathfrak{B}$ and all $u \in \mathfrak{U}(e)$. Then $A(x, a, u) = 0$ for $x \in \mathfrak{B}$, $a \in \mathfrak{A}$, $u \in \mathfrak{U}$. Now let $a, b \in \mathfrak{A}$, $u, v \in \mathfrak{U}$. Then

$$(au)(bv) = ((au)b)v = ((ua)b)v = (u(ab))v = ((ab)u)v = (ab)(uv).$$

This completes the proof.

LEMMA 2. *Let $\mathfrak{A}$ and $\mathfrak{U}$ be alternative algebras with identities. Then $\mathfrak{B} = \mathfrak{A} \otimes \mathfrak{U}$ is alternative if and only if either $\mathfrak{A}$ and $\mathfrak{U}$ are associative or one of these algebras is associative and commutative.*

Proof. If $a, b, c \in \mathfrak{A}$ and $u, v, w \in \mathfrak{U}$, then

$$\begin{aligned} A(a \otimes u, b \otimes v, c \otimes w) &= (ab)c \otimes (uv)w - a(bc) \otimes u(vw) \\ &= (ab)c \otimes (uv)w - (ab)c \otimes u(vw) \\ &\quad + (ab)c \otimes u(vw) - a(bc) \otimes u(vw) \\ &= (ab)c \otimes A(u, v, w) + A(a, b, c) \otimes u(vw). \end{aligned}$$

[3] A finite dimensional alternative algebra $\mathfrak{A}$ is called separable if $\mathfrak{A}_\Omega$ is semi-simple for Ω the algebraic closure of the base field.

If $\mathfrak{B}$ is alternative, then

$$0 = [ab]c \otimes A(u, v, w) + A(a, b, c) \otimes (u(vw) - v(uw))$$

where $[ab] = ab - ba$. We distinguish five cases:

(i) There exist u, v, w such that $A(u, v, w)$ and $u(vw) - v(uw)$ are linearly independent. Then $A(a, b, c) \equiv 0$ and $[ab]c \equiv 0$. This means that $\mathfrak{A}$ is associative and commutative.

(ii) There exist u, v, w such that $A(u, v, w) \neq 0$, $u(vw) - v(uw) \neq 0$ but these are linearly independent. Then $u(vw) - v(uw) = \lambda A(u, v, w)$, $\lambda \neq 0$. Hence $[ab]c = -\lambda A(a, b, c)$ for all a, b, c. Setting $c = 1$ gives $[ab] = 0$. Since $\lambda \neq 0$, $A(a, b, c) = 0$ and $\mathfrak{A}$ is associative and commutative.

(iii) There exist u, v, w such that $A(u, v, w) \neq 0$ but $u(vw) = v(uw)$. Then $[ab]c = 0$ and $[ab] = 0$. The original relation now gives $A(a, b, c) \equiv 0$ so that again $\mathfrak{A}$ is associative and communicative.

(iv) $A(u, v, w) \equiv 0$ but there is a u, v, w such that $u(vw) \neq v(uw)$. Then $A(a, b, c) = 0$ and both $\mathfrak{U}$ and $\mathfrak{A}$ are associative.

(v) $A(u, v, w) \equiv 0$ and $u(vw) \equiv v(uw)$. Here $\mathfrak{U}$ is associative and commutative.

We can now prove

THEOREM 1. *Let $\mathfrak{B}$ be an alternative algebra with an identity 1 and $\mathfrak{A}$ a subalgebra containing 1 and having the structure of a Cayley algebra. Then $\mathfrak{B} = \mathfrak{A} \otimes \mathfrak{U}$ where $\mathfrak{U}$ is the center[4] of $\mathfrak{B}$.*

Proof. We consider $\mathfrak{B}$ as an $\mathfrak{A}$-bimodule. It is known that every $\mathfrak{A}$-bimodule in which 1 is the identity operator (left and right) is a direct sum of bimodules isomorphic to $\mathfrak{A}$. This implies that $\mathfrak{B}$ contains a set of elements $\{u_\alpha\}$ (the images of 1 in the isomorphism with $\mathfrak{A}$) such that $1u_\alpha = u_\alpha = u_\alpha 1$, $au_\alpha = u_\alpha a$, $a(bu_\alpha) = (ab)u_\alpha$ for all $a, b \in \mathfrak{A}$ and every element of $\mathfrak{B}$ can be written in one and only one way in the form $\sum a_\alpha u_\alpha$, $a_\alpha \in \mathfrak{A}$. Let $\mathfrak{U}$ be the subset of elements of $\mathfrak{B}$ which commute with every $a \in \mathfrak{A}$. By Lemma 1, $\mathfrak{U}$ is a subalgebra. Evidently the $u_\alpha \in \mathfrak{U}$. On the other hand, let $x = \sum a_\alpha u_\alpha \in \mathfrak{U}$, $a_\alpha \in \mathfrak{A}$. Then Lemma 1 shows that $\sum (aa_\alpha)u_\alpha = \sum (a_\alpha a)u_\alpha$. Hence $a_\alpha a = aa_\alpha$ for all a. Since $\mathfrak{A}$ is central this means that $a_\alpha = \rho_\alpha 1$, ρ_α in the base field.

[4] The center of a non-associative algebra is the set of elements which commute with every element and associate with every pair of elements. The center is a commutative and associative subalgebra.

Hence $\mathfrak{U}$ is the space spanned by the u_α. It is now clear that $\mathfrak{B} = \mathfrak{A} \otimes \mathfrak{U}$. Lemma 2 then shows that $\mathfrak{U}$ is associative and commutative. It follows that $\mathfrak{U}$ is contained in the center. On the other hand, it is clear from the definition that $\mathfrak{U}$ contains the center.

The assumption that the identity of $\mathfrak{A}$ is an identity in $\mathfrak{B}$ can be dropped. One then obtains the following extension of Theorem 1:

THEOREM 2. *Let $\mathfrak{B}$ be an alternative algebra and $\mathfrak{A}$ a Cayley subalgebra. Then $\mathfrak{B} = \mathfrak{B}_1 \oplus \mathfrak{B}_0$ where the $\mathfrak{B}_i$ are ideals and $\mathfrak{B}_1 \supseteq \mathfrak{A}$ and is a Kronecker product $\mathfrak{B}_1 = \mathfrak{A} \otimes \mathfrak{U}$, $\mathfrak{U}$ the center of $\mathfrak{B}_1$.*

Proof. Let 1 be the identity of $\mathfrak{A}$ and let $\mathfrak{B} = \mathfrak{B}_{00} + \mathfrak{B}_{01} + \mathfrak{B}_{10} + \mathfrak{B}_{11}$ be the Peirce decomposition relative to 1. Thus

$$\mathfrak{B}_{ij} = \{b_{ij} \mid 1b_{ij} = \delta_{ij}e_{ij},\ b_{ij}1 = \delta_{ji}b_{ij}\}.$$

It is known that $\mathfrak{B}_{01}\mathfrak{A} \subseteq \mathfrak{B}_{01}$, $\mathfrak{A}\mathfrak{B}_{01} \subseteq \mathfrak{B}_{01}$ and if $a_1a' \in \mathfrak{A}$, $a_{01} \in \mathfrak{B}_{01}$ then $(aa')b_{01} = a(a'b_{01})$ (cf. [2], § 6). Since $\mathfrak{A}$ is not associative, this implies that $\mathfrak{B}_{01} = 0$. Similarly $\mathfrak{B}_{10} = 0$. Hence $\mathfrak{B} = \mathfrak{B}_1 \oplus \mathfrak{B}_0$ where $\mathfrak{B}_i \equiv \mathfrak{B}_{ii}$ is an ideal. Since $1 \in \mathfrak{B}_1$, $\mathfrak{A} \subseteq \mathfrak{B}_1$ and $\mathfrak{B}_1 = \mathfrak{A} \otimes \mathfrak{U}$, $\mathfrak{U}$ the center of $\mathfrak{B}_1$.

2. The Jordan theorem. For the proof of the factorization theorem for Jordan algebras, the following three lemmas will be required.

LEMMA 1. *Let $\mathfrak{R}$ be a Jordan algebra containing an identity 1 and elements e_{ii}, $u_{ij} = u_{ji}$, $i \neq j$, $i, j = 1, 2, \cdots, n$ such that*

$$(1) \qquad e^2_{ii} = 2e_{ii},\ e_{ii}u_{ij} = u_{ij},\ u^2_{ij} = 2(e_{ii} + e_{jj})\ u_{ij}u_{jk} = u_{ik},\quad i, j, k \neq$$

and all other products are 0. Assume, moreover, that $\frac{1}{2}\sum e_{ii} = 1$ and $n \geqq 3$. Then $\mathfrak{R}$ is isomorphic to the subalgebra $\mathfrak{H}$ of hermitian matrices of a suitable matrix algebra $\mathfrak{D}_{nJ}$ where $\mathfrak{D}$ is an involutorial alternative algebra which is associative if $n > 3$ and, in any case, has the property that its self-adjoint elements are in the nucleus.

We recall that the *nucleus* of an algebra is the subset of elements n such that

$$A(a, b, n) = A(a, n, b) = A(n, a, b) = 0 \text{ for all } a, b.$$

Lemma 1 is proved in a forthcoming paper ([2], Th. 9.1) by the author. We remark also that in the identification of $\mathfrak{R}$ with $\mathfrak{H}$ we may identify e_{ii} with (i, i)-matrix unit and u_{ij} with $e_{ij} + e_{ji}$, e_{ij} the (i, j)-matrix unit. We shall also require two other results on Jordan algebras ([2], Th. 7.1 and Th. 7.3) as follow.

LEMMA 2. *Let $\mathfrak{J}$ be a subalgebra of the algebra $\mathfrak{R}(=\mathfrak{H})$ of the preceding lemma containing the elements e_{ii}, u_{ij}. Then $\mathfrak{J}=\mathfrak{R}\cap\mathfrak{E}_n$ where $\mathfrak{E}$ is a subalgebra of $\mathfrak{D}$ invariant under the involution in $\mathfrak{D}$.*

LEMMA 3. *Let $\mathfrak{R}_1$ and $\mathfrak{R}_2$ be two algebras of the type given in Lemma 1 with the same n for the two. Let S be a homomorphism of $\mathfrak{R}_1$ onto $\mathfrak{R}_2$ mapping $e_{ii}^{(1)}$, $u_{ij}^{(1)}$ onto the corresponding $e_{ii}^{(2)}$, $u_{ij}^{(2)}$. Then S can be extended to a homomorphism of $\mathfrak{D}_{1n}$ onto $\mathfrak{D}_{2n}$.*

If S is an isomorphism, it follows that its extension is an isomorphism. In any case the extension maps the system of matrix units of $\mathfrak{D}_{1n}$ onto that of $\mathfrak{D}_{2n}$. It follows that if S is an isomorphism then $\mathfrak{D}_1$ and $\mathfrak{D}_2$ are isomorphic.

We can now prove

THEOREM 3. *Let $\mathfrak{J}$ be an exceptional simple Jordan algebra of finite dimensions and let $\mathfrak{R}$ be an arbitrary Jordan algebra containing $\mathfrak{J}$ and having the same identity as $\mathfrak{R}$. Then $\mathfrak{R}$ is a Kronecker product $\mathfrak{R}=\mathfrak{J}\otimes\mathfrak{C}$ where $\mathfrak{C}$ is the center of $\mathfrak{R}$.*

Proof. The argument used in the proof of Theorem 1 reduces the consideration to the case of an algebraically closed base field. In this case we may assume that the exceptional simple Jordan algebra is the algebra of hermitian Cayley matrices. Hence $\mathfrak{J}$ contains a set of elements e_{ii}, u_{ij}, $i,j=1,2,3$ satisfying the conditions of Lemma 1. Since $\mathfrak{R}\supseteq\mathfrak{J}$ and has the same identity as $\mathfrak{J}$, $\mathfrak{R}$ can be identified with the subalgebra of hermitian matrices in an algebra $\mathfrak{D}_3$ where $\mathfrak{D}$ is an involutorial alternative algebra whose self-adjoint elements are in the nucleus. We may suppose also that the elements e_{ii} are the matrix units e_{ii} and $u_{ij}=e_{ij}+e_{ji}$, $i\neq j$. Then by Lemma 2, $\mathfrak{J}=\mathfrak{R}\cap\mathfrak{E}_3$ where $\mathfrak{E}$ is a subalgebra of $\mathfrak{D}$ closed under the involution. By Lemma 3, $\mathfrak{E}$ is a Cayley algebra. Then, by Theorem 1, $\mathfrak{D}=\mathfrak{E}\otimes\mathfrak{C}$ where $\mathfrak{C}$ is the center of $\mathfrak{D}$. Let u be a skew element of $\mathfrak{C}$ ($\bar{u}=-u$) and let c be a skew element $\neq 0$ in $\mathfrak{E}$. (Elements c exist since the nucleus of $\mathfrak{E}$ is the set of multiples of 1.) If $u\neq 0$, cu is a self-adjoint element not contained in the nucleus of $\mathfrak{D}$. Hence $u=0$ and $\mathfrak{C}$ contains only self-adjoint elements. Thus $\mathfrak{C}\subseteq\mathfrak{R}$. It is now immediate that $\mathfrak{C}$ is the center of $\mathfrak{R}$ and that $\mathfrak{R}=\mathfrak{J}\otimes\mathfrak{C}$.

As in the alternative case we have the following extension of Theorem 3.

THEOREM 4. *Let $\mathfrak{J}$ be an exceptional simple Jordan algebra and let $\mathfrak{R}$ be an arbitrary Jordan algebra containing $\mathfrak{J}$ as a subalgebra. Then $\mathfrak{R}$ is a direct sum of two ideals $\mathfrak{R}=\mathfrak{R}_0\oplus\mathfrak{R}_1$ where $\mathfrak{R}_1\supseteq\mathfrak{J}$ and $\mathfrak{R}_1=\mathfrak{J}\otimes\mathfrak{C}$, $\mathfrak{C}$ the center of $\mathfrak{R}_1$.*

Proof. Let 1 be the identity of $\mathfrak{J}$ and let $\mathfrak{R} = \mathfrak{R}_1 \oplus \mathfrak{R}_{\frac{1}{2}} \oplus \mathfrak{R}_0$ be the Peirce decomposition of $\mathfrak{R}$ relative to the idempotent element 1. Then $\mathfrak{R}_1$ is a subalgebra containing $\mathfrak{J}$ and having the same identity. Hence $\mathfrak{R}_1 = \mathfrak{J} \otimes \mathfrak{C}$, $\mathfrak{C}$ the center of $\mathfrak{R}_1$. $\mathfrak{J}$ maps $\mathfrak{R}_{\frac{1}{2}}$ into itself and the correspondence $j \to S_j$ where $j \in \mathfrak{J}$ and S_j denotes the multiplication $x \to xj$, $x \in \mathfrak{R}_{\frac{1}{2}}$ is a special representation of $\mathfrak{J}$ ([2], § 6). Since $\mathfrak{J}$ is simple and exceptional it follows that $S_j = 0$ for all j. For $j = 1$ this gives $\mathfrak{R}_{\frac{1}{2}} = 0$. Hence $\mathfrak{R} = \mathfrak{R}_1 \oplus \mathfrak{R}_0$. Since $\mathfrak{R}_1\mathfrak{R}_0 = 0$ and $\mathfrak{R}_0^2 \subseteq \mathfrak{R}_0$ this proves the theorem.

Concluding remarks. We have recently proved the converse of Wedderburn's theorem: If $\mathfrak{A}$ is an associative algebra with an identity which is a Kronecker factor of every containing associative algebra having the same identity, then $\mathfrak{A}$ is finite dimensional central simple.[5] It can be seen from the representation theory of [2] that no special central simple Jordan algebra has the property of the theorem proved here. One might therefore conjecture that this property characterizes the exceptional simple Jordan algebras. In the alternative case one can prove, using Zorn's methods, that if $\mathfrak{B}$ is alternative with an identity and $\mathfrak{A}$ is an associative central simple subalgebra containing 1 and of more than 4 dimensions, then $\mathfrak{B}$ is necessarily associative. In this case Wedderburn's theorem applies. The example of the Cayley algebra shows that this is not the case if $\mathfrak{A}$ is a quaternion algebra. Nevertheless, it is conceivable that the structure of alternative algebras $\mathfrak{B}$ with 1 containing a quaternion subalgebra containing 1 can be fully determined.

YALE UNIVERSITY.

REFERENCES.

[1] A. A. Albert, "On simple alternative rings," *Canadian Journal of Mathematics*, vol. 2 (1952), pp. 129-135.

[2] N. Jacobson, "Structure of alternative and Jordan bimodules," forthcoming.

[3] I. Kaplansky, "Semi-simple alternative rings," *Portugaliae Mathematica*, vol. 10 (1951), pp. 37-50.

[4] R. D. Schafer, "The exceptional simple Jordan algebras," *American Journal of Mathematics*, vol. 70 (1948), pp. 82-84.

[5] T. Nakayama, "Wedderburn's theorem, weakly normal rings," *Journal of the Mathematical Society of Japan*, vol. 5 (1953), pp. 154-170.

[5] This will appear in a forthcoming book on structure theory of rings. See also [5], p. 169.

THE JACOBIAN VARIETY OF AN ALGEBRAIC CURVE.*†

By WEI-LIANG CHOW.

1. Preliminary remarks. We begin with a few remarks concerning terminology and notations. We shall follow in general the terminology and notations of A. Weil, as developed in his books [8], [9], with the modification that we shall operate throughout with points and (complete) varieties in projective spaces only. The projective space of dimension m will be denoted by $\mathfrak{S}_m$. If $(x) = (x_0, x_1, \cdots, x_m)$ is a point in $\mathfrak{S}_m$, the x_i being its homogeneous coordinates, and if K is a field, we shall denote by $K((x))$ the field generated over K by the adjunction of the mutual ratios of the x_i; similarly, if $(x), (y), \cdots$, are points in projective spaces, we denote by $K((x), (y), \cdots)$ the field generated over K by the adjunction of the mutual ratios of the homogeneous coordinates of every one of the points. The dimension of a point (x) over K is then the degree of transcendency of $K((x))$ over K, and (x) is said to be algebraic or rational over K if $K((x))$ is algebraic over K or coincides with K respectively; we shall also say that a homogeneous polynomial or form is rational over K, if the mutual ratios of its coefficients are all in K. We shall use the capital letters U, V, W, X, Y, Z (and these only) to denote indeterminates; for convenience, a system consisting of $m+1$ indeterminates $(X) = (X_0, X_1, \cdots, X_m)$ will be called a system of indeterminates in $\mathfrak{S}_m$.

We shall assume that the reader is familiar with the theory of associated forms of positive cycles in a projective space, as developed in Chow-van der Waerden [4]. For a positive cycle $\mathfrak{Z}$ of dimension r and degree d in a projective space $\mathfrak{S}_m$, the associated form is a form $f(V^0, \cdots, V^r)$ of degree d in each system of indeterminates $(V^i) = (V^i{}_0, V^i{}_1, \cdots, V^i{}_m)$, $i = 0, 1, \cdots, r$; this form is characterized by the following property: For any set of r hyperplanes in $\mathfrak{S}_m$ given by the equations $\sum_{j=0}^{m} v^i{}_j X_j = 0$, $i = 1, \cdots, r$, the form

* Received June 29, 1953.

† This paper was written in 1948; a brief summary of the results contained in it appeared under the same title in the Abstract of the addresses given at the Conference on Algebraic Geometry and Algebraic Number Theory, University of Chicago, 1949, pp. 25-27. It is published here with some minor revisions, mainly in the first two sections.

$f(V^0, v^1, \cdots, v^r)$ does not vanish identically if and only if $\mathfrak{Z}$ has a finite number of common intersections with these r hyperplanes, and when such is the case, the form $f(V^0, v^1, \cdots, v^r)$ dissolves into a product of d linear forms, giving all the common intersections with the proper multiplicities. If we regard the set of coefficients (f) of the form $f(V^0, \cdots, V^r)$, arranged in an arbitrary but fixed order, as a point in a projective space $\mathfrak{S}_t$, and if (f') is a specialization of (f) over any field K, then it follows from the main theorem of associated form ([4], Satz 2) that the point (f') corresponds also to the associated form $f'(V^0, \cdots, V^r)$ of a positive cycle $\mathfrak{Z}'$ of dimension r and degree d in $\mathfrak{S}_m$; this positive cycle $\mathfrak{Z}'$ is said to be a specialization of $\mathfrak{Z}$ over K. A property or relation concerning a point (x) in $\mathfrak{S}_m$ and a point (y) in $\mathfrak{S}_n$ is said to be algebraic over a field K, if there is a set of forms $F_j(X, Y)$ in $K[X, Y]$, homogeneous in each of the systems of indeterminates $(X) = (X_0, X_1, \cdots, X_m)$ and $(Y) = (Y_0, Y_1, \cdots, Y_n)$, such that the property holds for any two points (x) and (y) in $\mathfrak{S}_m$ and $\mathfrak{S}_n$ respectively if and only if the equations $F_j(x, y) = 0$ hold; we then say that the property in question can be expressed as an algebraic condition (over K) in terms of the points (x) and (y). In case the set $F_j(X, Y)$ contains forms in (X) or (Y) alone, the point (x) or (y) can be restricted to the points of a certain bunch of varieties in $\mathfrak{S}_m$ or $\mathfrak{S}_n$ respectively; in such a case the property in question is often defined only for the points of certain bunches of varieties in $\mathfrak{S}_m$ and $\mathfrak{S}_n$, so that the equations defining these bunches of varieties are already implicitly contained in the definition of the property. The following fact is important for us later: The property that a given algebraic relation over K holds for a point (y) in $\mathfrak{S}_n$ and *every* point (x) in a positive cycle $\mathfrak{Z}$ in $\mathfrak{S}_m$ can itself be expressed as an algebraic condition over K in terms of (y) and the point (f) in $\mathfrak{S}_t$. To show this, we refer to the proof of Satz 2 in [4] and consider the equations (2), (3), (5) listed there; in our present notations these can be writen as a set of equations

$$G_j(V^1, \cdots, V^r, X^1, \cdots, X^d, Z) = 0,$$

where $(X^1), \cdots, (X^d)$ are d distinct systems of indeterminates in $\mathfrak{S}_m$ and (Z) is a system of indeterminates in $\mathfrak{S}_t$. Let $R_j(V^1, \cdots, V^r, Y, Z) = 0$ be the set of equations which we obtain by eliminating the indeterminates $(X^1), \cdots, (X^d)$ from the set of equations $F_j(X^1, Y) = 0, \cdots, F_j(X^d, Y) = 0$, $G_j(V^1, \cdots, V^r, X^1, \cdots, X^d, Z) = 0$, and let $A_j(Y, Z)$ be the set of all coefficients of the forms $R_j(V^1, \cdots, V^r, Y, Z)$, considered as forms in $(V^1), \cdots, (V^r)$ only; then it is easily seen from the cited proof in [4]

that the property in question holds for (y) and $\mathfrak{Z}$ if and only if $A_j(y, f) = 0$. It is clear that this result remains true if more than one point (y) or more than one cycle $\mathfrak{Z}$ are involved, and it is also clear that it remains true even if the original algebraic property does not actually involve the point (y) at all, so that it is simply an algebraic property for the points of the positive cycle or cycles. For later reference we shall state our result as a lemma:

LEMMA 1. *Let $\mathfrak{Z}^1, \cdots, \mathfrak{Z}^s$ be positive cycles in projective spaces and let $(f^1), \cdots, (f^s)$ be the systems of coefficients of the associated forms of $\mathfrak{Z}^1, \cdots, \mathfrak{Z}^s$ respectively; let $(y^1), \cdots, (y^h)$ be points in projective spaces. Then the property that a given algebraic relation over a field K holds for the system of points $(y^1), \cdots, (y^h)$ and every system of points $(x^1), \cdots, (x^s)$, where each (x^i) runs through all the points in $\mathfrak{Z}^i$, can be expressed as an algebraic condition over K in terms of $(y^1), \cdots, (y^h)$ and $(f^1), \cdots, (f^s)$.*

It is easily seen from the definition of the associated form that if the positive cycle $\mathfrak{Z}$ is rational over a field K, then its associated form $f(V^0, \cdots, V^r)$ is also rational over K, i. e. $K((f)) = K$. We shall now show that the converse of this statement is also true, provided that $\mathfrak{Z}$ contains no component with a multiplicity which is divisible by the characteristic of the universal domain; more precisely, we shall now prove the following lemma:

LEMMA 2. *Let $\mathfrak{Z}$ be a positive cycle in a projective space and let (f) be the set of coefficients of the associated form of $\mathfrak{Z}$; if $\mathfrak{Z}$ contains no component with a multiplicity which is divisible by the characteristic of the field, then $\mathfrak{Z}$ is rational over the fields $k_0((f))$, where k_0 is the prime field.*

Proof. Without any loss of generality we can assume that the form $f(V^0, \cdots, V^r)$ is irreducible over the field $K_0 = k_0((f))$, for otherwise we can apply the same argument to each irreducible factor of $f(V^0, \cdots, V^r)$ over K_0; then our assumption about the multiplicities of the components in $\mathfrak{Z}$ implies that the form $f(V^0, \cdots, V^r)$ has no multiple factors over $\bar{K}_0$. It follows that $f(V^0, \cdots, V^r)$ is the product of a number s of absolutely irreducible forms, all distinct and conjugate to each other over K_0, and each one of these forms is rational over a finite extension of degree s over K_0. In order to prove our assertion, it is therefore sufficient to restrict ourselves to the case where the form $f(V^0, \cdots, V^r)$ is absolutely irreducible and the positive cycle $\mathfrak{Z}$ is a variety defined over a field K containing K_0, and to show in this case the variety $\mathfrak{Z}$ is also defined over K_0. Consider the equation $f(V^0, \cdots, V^r) = 0$; since the form $f(V^0, \cdots, V^r)$ is absolutely irreducible,

this equation defines a variety $\mathfrak{T}$ of dimension $(r+1)m-1$ in the $(r+1)$-fold product space $S_m \times \cdots, \times S_m$, defined over the field K_0. Let (x) be a generic point of $\mathfrak{Z}$ over K, and let $(v^0), \cdots, (v^r)$ be $r+1$ independent generic solutions of the linear equation $\sum_{j=0}^{m} x_j V_j = 0$ over the field $K((x))$; if r' $(\geq r)$ is the dimension of (x) over K_0, then the point $(v^0, \cdots, v^r)$ in $\mathfrak{T}$ has the dimension

$$(r+1)(m-1)+r' = (r+1)m-1+(r'-r) \geq (r+1)m-1$$

over K_0, from which it follows that $r=r'$ and that $(v^0, \cdots, v^r)$ is a generic point of $\mathfrak{T}$ over K_0. Since the point $(v^0, \cdots, v^r)$ has the same dimension $(r+1)m-1$ over both K and K_0, the fields $K_0((v^0), \cdots, (v^r))$ and K must be independent with respect to each other over K_0, and since $K_0((v^0), \cdots, (v^r))$ is a regular extension of K_0, this means that $K_0((v^0), \cdots, (v^r))$ and K are linearly disjoint with respect to each other over K_0. Without any loss of generality, we can assume that $x_0 \neq 0$ and that $x_m = 0$ in case any one of the coordinates x_i vanishes; then we can take all $v^i{}_j$ $(j \neq 0)$ to be independent variables over $K((x))$ and we have $v^i{}_m \neq 0$ for all $i = 0, 1, \cdots, r$. We have then the relations $\sum_{j=1}^{m} (v^i{}_j / v^i{}_m)(x_j/x_0) = -v^i{}_0/v^i{}_m$, $i = 0, 1, \cdots, r$; the first r equations show ([8], Chapter V, Theorem 1) that (x) is separably algebraic over $K_0((v^0), \cdots, (v^{r-1}))$, and then the last equation shows that (x) is rational over $K_0((v^0), \cdots, (v^r))$. Since $K_0((x))$ is subfield in $K_0((v^0), \cdots, (v^r))$, it is also linearly disjoint with respect to K over K_0; this shows that the variety $\mathfrak{Z}$ is also the locus of (x) over K_0 and hence is defined over K_0.

2. The problem. Let $\mathfrak{C}$ be a curve in the projective space $\mathfrak{S}_m$ of dimension m, defined over a field k, and let h and g be the degree and genus of $\mathfrak{C}$ respectively. A positive divisor $\mathfrak{p}$ of degree n in $\mathfrak{C}$ can be considered as a positive 0-cycle in $\mathfrak{S}_m$ and as such it can be represented by its associated form $p(V)$. Let $\Omega_0(V), \Omega_1(V), \cdots, \Omega_t(V)$ be the set of all power products of degree n in (V), arranged in an order which is arbitrary but fixed once for all; if we set $p(V) = \sum_{j=0}^{t} p_j \Omega_j(V)$, then the set of coefficients $(p) = (p_0, p_1, \cdots, p_t)$, arranged in this fixed order, determines a point in the projective space $\mathfrak{S}_t$ of dimension t, so that every positive divisor of degree n in $\mathfrak{C}$ is represented in a one-to-one manner by a point in $\mathfrak{S}_t$. We shall assume that the curve $\mathfrak{C}$ is free of singularities, so that every 0-cycle in $\mathfrak{S}_m$

which is contained in $\mathfrak{C}$ can also be considered as a divisor in $\mathfrak{C}$; then the set of all points in $\mathfrak{S}_t$ which represent positive divisors of degree n in $\mathfrak{C}$ is a variety $\mathfrak{C}^n$ of dimension n, defined over k. In fact, let $\mathfrak{y}$ be the positive divisor of degree n consisting of n independent generic points $(x^1), \cdots, (x^n)$ of $\mathfrak{C}$ over k, and let (y) be the point in $\mathfrak{S}_t$ representing $\mathfrak{y}$; then, for any positive divisor $\mathfrak{p}$ of degree n in $\mathfrak{C}$, the point (p) is a specialization of the point (y) over k, and conversely, by the main theorem of associated form, every such specialization (p) determines a positive 0-cycle of degree n in $\mathfrak{S}_m$ which is contained in $\mathfrak{C}$ and hence is a positive divisor of degree n in $\mathfrak{C}$. Since $k((y))$ is contained in the regular extension $k((x^1), \cdots, (x^n))$ of k, it is also a regular extension of k, and since $k((x^1), \cdots, (x^n))$ has the degree of transcendency n over k and is an algebraic extension of $k((y))$, the field $k((y))$ must have also the degree of transcendency n over k. It follows that the point (y) has a locus over k which is a variety of dimension n and which evidently coincides with $\mathfrak{C}^n$. As we shall see later, this variety $\mathfrak{C}^n$ is also free of singularities, but we shall not need this fact for the present. It is clear that if a positive divisor $\mathfrak{p}$ is rational over a field K, then the point (p) representing it is also rational over K; the converse of this statement is also true, as we have shown in [3], so that for any point (p) in $\mathfrak{C}^n$ the positive divisor $\mathfrak{p}$ determined by it is rational over $k((p))$.

Two divisors $\mathfrak{p}$, $\mathfrak{q}$ in $\mathfrak{C}$ are said to be linearly equivalent and we shall write $\mathfrak{p} \sim \mathfrak{q}$, if $\mathfrak{p} - \mathfrak{q}$ is the divisor of a rational function on $\mathfrak{C}$, or in other words, if $\mathfrak{p} - \mathfrak{q}$ is a principal divisor in $\mathfrak{C}$. It is well known that the set of all principal divisors in $\mathfrak{C}$ forms a subgroup $\mathfrak{G}_l(\mathfrak{C})$ in the group $\mathfrak{G}(\mathfrak{C})$ of all divisors in $\mathfrak{C}$, and that the quotient group of $\mathfrak{G}(\mathfrak{C})$ over $\mathfrak{G}_l(\mathfrak{C})$ is the group $\mathfrak{D}$ of divisor classes in $\mathfrak{C}$, each class consisting of all divisors which are linearly equivalent to each other. In this group $\mathfrak{D}$ the classes of degree zero, sometimes called the zero classes, again form a subgroup $\mathfrak{D}_0$, while the classes of any one fixed degree n form a coset $\mathfrak{D}_n$ of $\mathfrak{D}_0$ in $\mathfrak{D}$; if we denote by $\mathfrak{G}_a(\mathfrak{C})$ the group of all divisors of degree zero in $\mathfrak{C}$, then $\mathfrak{D}_0$ is the quotient group of $\mathfrak{G}_a(\mathfrak{C})$ over $\mathfrak{G}_l(\mathfrak{C})$. If we let any one fixed divisor class $\mathfrak{O}$ of $\mathfrak{D}_n$ correspond to the principal class of $\mathfrak{D}_0$, then the relation $\mathfrak{D}_n = \mathfrak{O} + \mathfrak{D}_0$ establishes a one-to-one correspondence between the classes of $\mathfrak{D}_n$ and those of $\mathfrak{D}_0$. Through this correspondence the group addition of $\mathfrak{D}_0$ will induce in $\mathfrak{D}_n$ the following operation: Given any two classes $\mathfrak{P}, \mathfrak{Q}$ in $\mathfrak{D}_n$, the class $\mathfrak{P} + \mathfrak{Q} - \mathfrak{O}$, also belonging to $\mathfrak{D}_n$, is called the sum of $\mathfrak{P}$ and $\mathfrak{Q}$. We shall call this operation in $\mathfrak{D}_n$ also addition (with respect to the class $\mathfrak{O}$).

A divisor class $\mathfrak{P}$ is said to be rational over a field K, if it contains a

divisor $\mathfrak{p}$ which is rational over K. It is well known that the set of positive divisors in the class $\mathfrak{P}$ constitutes a linear system; that is, there exists a linear system of hypersurfaces in $\mathfrak{S}_m$ which cut out on $\mathfrak{C}$, besides a fixed component, exactly the positive divisors of the class $\mathfrak{P}$. If the class $\mathfrak{P}$ is rational over K, then ([8], Chapter VIII, Theorem 10) this linear system of hypersurfaces has a basis consisting of hypersurfaces which are defined over K. We recall that according to the Theorem of Riemann-Roch, the linear system of positive divisors of a class of degree $n > 2g - 2$ has exactly the dimension $n - g$.

A variety $\mathfrak{J}$, defined over a field K containing k, is called a Jacobian variety of the curve $\mathfrak{C}$, if it has the following properties: (1) $\mathfrak{J}$ is an Abelian variety defined over K; (2) there is a homomorphic mapping Φ of $\mathfrak{G}_a(\mathfrak{C})$ onto $\mathfrak{J}$, called the canonical homomorphism, whose kernel is precisely $\mathfrak{G}_1(\mathfrak{C})$; (3) the homomorphism Φ is rational and is defined over K in the following sense: if $\mathfrak{p}_u$ is a rational divisor over $K(u)$, where (u) is any set of elements (in the universal domain), then the point $\Phi(\mathfrak{p}_u)$ is also rational over $K(u)$, and if $\mathfrak{p}_\mu$ is any specialization of $\mathfrak{p}_u$ over a specialization $(u) \to (\mu)$ over K, then $\Phi(\mathfrak{p}_\mu)$ is also a specialization of $\Phi(\mathfrak{p}_u)$ over the specialization $(u) \to (\mu)$ over K; (4) the rational homomorphism Φ has the "universal maping" property: if Ψ is any rational homomorphism of $\mathfrak{G}_a(\mathfrak{C})$ into any Abelian variety $\mathfrak{A}$, then Ψ is the product of Φ and a rational homomorphism of $\mathfrak{J}$ into $\mathfrak{A}$. It can be easily seen from the "universal mapping" property (4) that the Jacobian variety $\mathfrak{J}$, it it exists, is uniquely determined up to a birational isomorphism, so that we can sometimes speak of *the* Jacobian variety. However, the question whether such a Jacobian variety exists at all is not a simple one, and this existence problem of the Jacobian variety is one of principal importance not only for the theory of algebraic curves, but also for the theory of Abelian varieties. In this paper we shall offer a general solution of this problem.

In the classical case, where the ground field is the field of complex numbers, a proof of the existence of the Jacobian variety can be obtained from the known results in the literature in the following way. (The terminology and notations in this one paragraph do not follow strictly the conventions set forth in section 1.) Let $I_1, \cdots, I_g$ be a system of g independent Abelian differentials of the first kind on $\mathfrak{C}$, and let $\gamma_1, \cdots, \gamma_{2g}$ be a base of the first Betti group of $\mathfrak{C}$ and we set $p_{ij} = \int_{\gamma_i} I_j$; let $\mathfrak{E}_g$ be the complex linear space of dimension g, considered as an analytic (= complex-analytic) group variety, and let $[P]$ be the discrete subgroup generated by

the $2g$ rows of the matrix $P = (p_{ij})$, each row being considered as a point in E_g, and let $\mathfrak{E}_g/[P]$ be the quotient group variety of $\mathfrak{E}_g$ over $[P]$. Let (x_0) be any fixed point in $\mathfrak{C}$ and let ϕ be the analytic mapping of $\mathfrak{C}$ into $\mathfrak{E}_g/[P]$ defined by the formula $\phi(x) = (\int_{x_0}^{x} I_1, \cdots, \int_{x_0}^{x} I_g) \pmod{[P]}$; then, if $\mathfrak{p} = \sum n_i(x_i)$ is any divisor in $\mathfrak{G}_a(\mathfrak{C})$, the correspondence $\mathfrak{p} \to \sum n_i\phi(x_i)$ defines a homomorphism Φ of $\mathfrak{G}_a(\mathfrak{C})$ onto $\mathfrak{E}_g/[P]$, whose kernel is, by the Abel's Theorem, precisely the group $\mathfrak{G}_1(\mathfrak{C})$. If Ψ is any homomorphism of $\mathfrak{G}_a(\mathfrak{C})$ into an analytic commutative group variety $\mathfrak{A}$, then $\psi(x) = \Psi((x) - (x_0))$ is a mapping of $\mathfrak{C}$ into $\mathfrak{A}$ and we have $\Psi(\mathfrak{p}) = \sum n_i\psi(x_i)$; we shall say that the homomorphism Ψ is analytic if the mapping ψ is analytic, and we observe that, in case $\mathfrak{A}$ is algebraic, the homomorphism Ψ is rational if and only if ψ is rational (over the complex ground field). It is clear that the homomorphism Φ defined above is analytic; we shall show that if Ψ is analytic and if $\mathfrak{A}$ is analytically isomorphic to a complex torus, then Ψ is the product of Φ and an analytic homomorphism of $\mathfrak{E}_g/[P]$ into $\mathfrak{A}$. We set $\mathfrak{A} = \mathfrak{E}_r/[Q]$, where $Q = (q_{ij})$ is a $(2r, r)$-matrix such that the matrix $(Q, \bar{Q})$ is non-singular, and let $z_1, \cdots, z_r$ be the coordinates of a variable point in $\mathfrak{E}_r$, so that $dz_1, \cdots, dz_r$ can be considered as Abelian differentials of the first kind on $\mathfrak{E}_r/[Q]$; it is well known that a base $\delta_1, \cdots, \delta_{2r}$ for the first Betti group of $\mathfrak{E}_r/[Q]$ can be so chosen that we have $q_{ij} = \int_{\delta_i} dz_j$. Let $d\psi_1, \cdots, d\psi_r$ be the Abelian differentials of the first kind on $\mathfrak{C}$ induced by $dz_1, \cdots, dz_r$ respectively through the mapping ψ, and let $\psi(\gamma_1), \cdots, \psi(\gamma_{2g})$ be the image cycles of $\gamma_1, \cdots, \gamma_{2g}$ respectively under ψ; then there exist a (g, r)-matrix $M = (m_{ij})$ and an integral $(2g, 2r)$-matrix $N = (n_{ij})$ such that we have the relations $d\psi_i = \sum_{j=1}^{g} m_{ji} I_j$, $\psi(\gamma_i) = \sum_{j=1}^{2r} n_{ij}\delta_j$, and

$$PM = \sum_{k=1}^{g} p_{ik}m_{kj} = \sum_{k=1}^{g} m_{kj} \int_{\gamma_i} I_k = \int_{\gamma_i} d\psi_j = \int_{\psi(\gamma_i)} dz_j = \sum_{k=1}^{2r} n_{ik} \int_{\delta_k} dz_j$$

$$= \sum_{k=1}^{2r} n_{ik}q_{kj} = NQ.$$

This shows that the linear mapping of $\mathfrak{E}_g$ into $\mathfrak{E}_r$ determined by the matrix M induces an analytic homomorphism Θ of $\mathfrak{E}_g/[P]$ into $\mathfrak{E}_r/[Q]$, and we have the relation

$$\psi(x) = (\int_{x_0}^{x} d\psi_1, \cdots, \int_{x_0}^{x} d\psi_r) = (\int_{x_0}^{x} I_1, \cdots, \int_{x_0}^{x} I_g)M \pmod{[Q]},$$

or $\psi(x) = \Theta\phi(x)$, from which it follows that $\Psi = \Theta\Phi$. Now, according to a result of Lefschetz ([5], p. 368), there exists an analytic isomorphism σ of $\mathfrak{E}_g/[P]$ onto an Abelian variety $\mathfrak{J}$; if we assume that there is an analytic isomorphism τ of $\mathfrak{E}_r/[Q]$ onto an Abelian variety $\mathfrak{A}$, and if we now write $\phi, \Phi, \psi, \Psi, \Theta$ in place of $\sigma\phi, \sigma\Phi, \tau\psi, \tau\Psi, \tau\Theta\sigma^{-1}$, respectively, then we have again the relations $\psi = \Theta\phi$ and $\Psi = \Theta\Phi$. Since ϕ, ψ, and Θ are now analytic mappings of an algebraic variety into another algebraic variety, it follows from a theorem of ours proved elsewhere ([2], Theorem VII) that they are all rational transformations, and hence the homomorphisms Φ and Ψ are also rational. This shows that $\mathfrak{J}$ is the Jacobian variety of $\mathfrak{C}$ with Φ as the canonical homomorphism.

In the general case of an arbitrary ground field, it is clear that the above highly transcendental proof cannot be carried over and that new methods of a purely algebraic nature have to be introduced. A. Weil, who first proposed and studied this problem, obtained a solution by a generalization of the concept of an algebraic variety ([9], § V); in fact, the attempt to solve this problem is one of the main reasons which led Weil to introduce his notation of an "abstract" variety, an algebraic variety in abstracto in contrast to the usual ones which are embedded in a projective space. As Weil has been able to extend most of the fundamental results in algebraic geometry to the "abstract" varieties, his "abstract" Jacobian variety proves to be almost just as useful as the projective Jacobian variety in the classical case and thus constitutes a satisfactory solution of the problem. Nevertheless, it remains an important and interesting question whether a Jacobian variety exists as a projective algebraic variety in the usual sense; furthermore, what is probably even more important from the algebraic viewpoint, there remains still unsolved the question of the field of definition for the Jacobian variety and the canonical homomorphism, as has been pointed out by Weil himself ([9], p. 68). Our solution of the problem will provide satisfactory answers to both these questions; we shall show that the Jacobian variety exists as a variety in a projective space, and that both the Jacobian variety and the canonical homomorphism are defined over the defining field of the curve $\mathfrak{C}$. This latter fact, which seems to be new even in the classical case, has important applications in the theory of Picard varieties over arbitrary ground fields, as will be shown in a forthcoming paper of ours.

We shall describe briefly the main idea underlying our construction of the Jacobian variety. Since the positive divisors in $\mathfrak{C}$ can be represented by the points of an algebraic variety, it is natural to try to represent the divisor classes by associating with them certain positive divisors. According

to the Riemann-Roch Theorem, each class of $\mathfrak{D}_g$ contains *in general* only one positive divisor; thus each point of $\mathfrak{C}^g$ will represent a class in $\mathfrak{D}_g$, and the representation is one-to-one except for the points of a proper subvariety in $\mathfrak{C}^g$. It is therefore natural that this variety $\mathfrak{C}^g$ has been usually taken as the starting point in the construction of the Jacobian variety; thus Weil's "abstract" Jacobian variety is essentially a collection of birational transforms of $\mathfrak{C}^g$ patched together in a suitable manner. Also van der Waerden, in a paper [7] which deals in reality with certain aspects of the Jacobian variety, has taken this variety $\mathfrak{C}^g$ as the starting point and has constructed by means of the associated forms a variety in a projective space, the points of which are in one-to-one correspondence with the classes in $\mathfrak{D}_g$; however, his results do not show that the so constructed variety is the Jacobian variety, as, for one thing, there is no proof that the variety is non-singular. We observe that in both cases the main difficulty lies in the existence of special divisor classes of degree g which are represented not by points but by subvarieties in $\mathfrak{C}^g$; one has then to cope with the fact that the divisor classes are represented by different types of geometrical entities. If now, instead of $\mathfrak{C}^g$, we consider the variety $\mathfrak{C}^n$ for sufficiently high n (say $n > 2g - 2$), then there will be no special divisor classes and every divisor class of degree n will be represented by a subvariety of dimension $n - g$ in $\mathfrak{C}^n$; the method of associated forms then will enable us to represent the divisor classes by ponts of a certain variety, and from the homogeneous nature of the construction one would expect the so constructed variety to be non-singular. This is the underlying idea of the method by which we shall construct the Jacobian variety in the next section; however, the proof that the so constructed variety is non-singular is not so simple and will be given later in section 5 as an application of a general theorem proved by us elsewhere. In the final section we then introduce the canonical homomorphism and complete the proof of the existence of the Jacobian variety.

3. Construction of the variety $\mathfrak{V}$. We start with the variety $\mathfrak{C}^n$ of n dimensions, defined over the field k, for a fixed $n > 2g - 2$. For any point (p) of $\mathfrak{C}^n$, such that the corresponding positive divisor $\mathfrak{p}$ is rational over a field K containing k, the Theorem of Riemann-Roch, applied to the curve $\mathfrak{C}$ over the field K, shows that the complete linear system of positive divisors on $\mathfrak{C}$ determined by the divisor $\mathfrak{p}$ has the dimension $n - g$ and is rational over the field K. In other words, the set of all positive divisors in the class $\mathfrak{P}$ determined by $\mathfrak{p}$ consists exactly of those divisors cut out on the cuve $\mathfrak{C}$, apart from a fixed component, by a linear system of $n - g + 1$ hyper-

surfaces $\sum_{j=0}^{n-g} U_j\psi_j(X) = 0$, where the $\psi_j(X)$ are forms of the same degree in $K[X]$ and the $(U) = (U_0, \cdots, U_{n-g})$ is a system of indeterminates; this implies also that no hypersurface of this system contains the curve $\mathfrak{C}$. Let $(u) = (u_0, \cdots, u_{n-g})$ be a set of independent variables over K, and let $\mathfrak{p}_u$ be the divisor cut out on $\mathfrak{C}$, apart from the fixed component, by the generic hypersurface $\sum_{j=0}^{n-g} u_j\psi_j(X) = 0$ of the system; then the divisor $\mathfrak{p}_u$ is rational over $K(u)$ and hence is represented by a rational point (p_u) over $K(u)$ in $\mathfrak{C}^n$. It is well known that over any specialization $(u) \to (\mu)$ over K, the divisor $\mathfrak{p}_u$ specializes into the divisor $\mathfrak{p}_\mu$ cut out on $\mathfrak{C}$, apart from the fixed component, by the hypersurface $\sum_{j=0}^{n-g} \mu_j\psi_j(X) = 0$. It follows that over any specialization $(u) \to (\mu)$ over K, there is a uniquely determined specialization (p_μ) of (p_u), which represents a divisor in $\mathfrak{P}$, and that in this way all positive divisors in $\mathfrak{P}$ can be obtained. Since $K((p_u))$, being a subfield of $K(u)$, is a regular extension of K, this shows that if we denote by $\mathfrak{G}_\mathfrak{P}$ the set of all points in $\mathfrak{C}^n$ which represent positive divisors in $\mathfrak{P}$, then $\mathfrak{G}_\mathfrak{P}$ is a variety of dimension $n - g$, defined over K, and the point (p_u) is a generic point of $\mathfrak{G}_\mathfrak{P}$ over K. Furthermore, the correspondence $u \to (p_u)$ defines a birational transformation of the projective space $\mathfrak{S}_{n-g}$ onto $\mathfrak{G}_\mathfrak{P}$, defined over K; in fact, the divisor $\mathfrak{p}_u$ consists of n distinct points $(\xi^1), \cdots, (\xi^n)$, which form a complete set of separable conjugates over $K((p_u))$, and the point (u) is determined rationally by the linear conditions $\sum_{j=0}^{n-g} u_j\psi_j(\xi^i) = 0$, $i = 1, \cdots, n$. We shall denote by $G_\mathfrak{P}(W) = G_\mathfrak{P}(W^0, \cdots, W^{n-g})$ the associated form of the variety $\mathfrak{G}_\mathfrak{P}$, where for each $i = 0, 1, \cdots, n - g$, the $(W^i) = (W^i{}_0, W^i{}_1, \cdots, W^i{}_t)$ is a system of indeterminates in $\mathfrak{S}_t$; since the variety $\mathfrak{G}_\mathfrak{P}$ is defined over K, the form $G_\mathfrak{P}(W)$ is rational over K.

Let (y) be a generic point of $\mathfrak{C}^n$ over k, and let $\mathfrak{y}$ be the corresponding positive divisor on $\mathfrak{C}$. Since $\mathfrak{y}$ is rational over $k((y))$ and consequently the class $\mathfrak{Y}$ determined by $\mathfrak{y}$ is also rational over $k((y))$, it follows that the associated form $G_\mathfrak{Y}(W)$ of the variety $\mathfrak{G}_\mathfrak{Y}$ is rational over $k((y))$. Let d be the degree of the variety $\mathfrak{G}_\mathfrak{Y}$ (as a subvariety in $\mathfrak{S}_t$), and let $\Delta_0(W)$, $\Delta_1(W), \cdots, \Delta_l(W)$ be the set of all power products of degree d in (W), arranged in an order which is arbitrary but fixed once for all; if we set $G_\mathfrak{Y}(W) = \sum_{j=0}^{l} G_j\Delta_j(W)$, then the set of coefficients $(G_\mathfrak{Y}) = (G_0, G_1, \cdots, G_l)$, arranged in this fixed order, determines a point in the projective space $\mathfrak{S}_l$

of dimension l, rational over $K((y))$. Since the variety $\mathfrak{C}^n$ is defined over k, it follows that the point $(G_\mathfrak{y})$ defines a variety in $\mathfrak{S}_l$ over k, which we shall denote by $\mathfrak{V}$; and the correspondence $(y) \to (G_\mathfrak{y})$ determines a rational transformation ϕ of $\mathfrak{C}^n$ onto $\mathfrak{V}$, defined over k. We maintain that the rational transformation has a unique value $\phi(p)$ at every point (p) in $\mathfrak{C}^n$, and that the positive cycle $\mathfrak{G}'_\mathfrak{P}$ of dimension $n-g$ and degree d in $\mathfrak{S}_t$ determined by the point $\phi(p)$ consists of exactly the one variety $\mathfrak{G}_\mathfrak{P}$, possibly with a certain multiplicity. To show this, we observe ([7], § 14) that the property that any two points (p) and (q) in $\mathfrak{C}^n$ represent linearly equivalent divisors is an algebraic condition over k; it follows then from this and Lemma 1 (section 1) that the property that a positive cycle $\mathfrak{G}$ of dimension $n-g$ and degree d in $\mathfrak{S}_t$ is contained in the variety $\mathfrak{G}_\mathfrak{P}$ can be expressed as an algebraic conidtion over k in terms of the point (p) in $\mathfrak{S}_t$ and the point (G) in $\mathfrak{S}_l$ representing $\mathfrak{G}$. In other words, there exists a set of forms $A_j(Y, Z)$ in $k[Y, Z]$, where the (Y) and (Z) are systems of indeterminates in $\mathfrak{S}_t$ and $\mathfrak{S}_l$ respectively, such that the cycle $\mathfrak{G}$ is contained in $\mathfrak{G}_\mathfrak{P}$ if and only if the equations $A_j(p, G) = 0$ hold. It is evident that the equations $A_j(y, G_\mathfrak{y}) = A_j(y, \phi(y)) = 0$ hold; if $\phi(p)$ is any specialization of $\phi(y)$ over the specialization $(y) \to (p)$ over k, then we have the relations $A_j(p, \phi(p)) = 0$. This shows that the positive cycle $\mathfrak{G}'_\mathfrak{P}$ determined by $\phi(p)$ can only have component varieties which are contained in $\mathfrak{G}_\mathfrak{P}$; and since every component in $\mathfrak{G}'_\mathfrak{P}$ has the same dimension $n-g$ as that of $\mathfrak{G}_\mathfrak{P}$, it must coincide with $\mathfrak{G}_\mathfrak{P}$, so that the positive cycle $\mathfrak{G}'_\mathfrak{P}$ is a multiple of the variety $\mathfrak{G}_\mathfrak{P}$ (we shall show in section 5 that $\mathfrak{G}'_\mathfrak{P}$ actually coincides with $\mathfrak{G}_\mathfrak{P}$). This evidently also shows that the specialization $\phi(p)$ is uniquely determined.

Thus we have shown that the points of the variety $\mathfrak{V}$ represent in a one-to-one manner the divisor classes of $\mathfrak{C}$, and that for any two divisors $\mathfrak{p}$ and $\mathfrak{q}$ in $\mathfrak{C}$, we have $\phi(p) = \phi(q)$ if and only if $\mathfrak{p} \sim \mathfrak{q}$. The next step is to define a composition function in $\mathfrak{V}$ which corresponds to the group addition in $\mathfrak{D}_0$, or rather the addition in $\mathfrak{D}_n$ with respect to some one fixed divisor class $\mathfrak{O}$. For this assume the integer n so chosen that there exists a divisor class $\mathfrak{O}$ in $\mathfrak{D}_n$ which is rational over k; it is easily seen that such an integer n exists. The corresponding point (G^0) on $\mathfrak{V}$ is then also rational over k. Let (G^1) and (G^2) be any two points in $\mathfrak{V}$, and let $\mathfrak{P}^1$ and $\mathfrak{P}^2$ be the divisor classes determined by (G^1) and (G^2) respectively; if (G^3) is the point in $\mathfrak{V}$ which represents the divisor class $\mathfrak{P}^3 = \mathfrak{P}^1 + \mathfrak{P}^2 - \mathfrak{O}$, then we define (G^3) as the sum of (G^1) and (G^2), and write $(G^3) = (G^1) + (G^2)$. We shall show that if (G^1) and (G^2) are two independent generic points of $\mathfrak{V}$ over k, then the point (G^3) is rational over the field $K = k((G^1), (G^2))$. We observe

first that since both (G^1) and (G^2) are generic points of $\mathfrak{V}$ over k, the positive cycles $\mathfrak{G}^1$ and $\mathfrak{G}^2$ determined by them are subvarieties in $\mathfrak{C}^n$, which by Lemma 2 (section 1), are defined over the fields $k((G^1))$ and $k((G^2))$ respectively, so that both are defined over K. Let (p^1) and (p^2) be independent generic points of the varieties $\mathfrak{G}^1$ and $\mathfrak{G}^2$ respectively over K; since the divisor class $\mathfrak{P}^1$ is rational over $K((p^1))$ and the divisor class $\mathfrak{P}^2$ is rational over $K((p^2))$, the class $\mathfrak{P}^3 = \mathfrak{P}^1 + \mathfrak{P}^2 - \mathfrak{D}$ is rational over the field $K((p^1), (p^2))$ and consequently we have $k((G^3)) \subset K((p^1), (p^2))$. Now, let (q^1) and (q^2) be another pair of such independent generic points of the varieties $\mathfrak{G}^1$ and $\mathfrak{G}^2$ respectively, both independent with respect to the field $K((p^1), (p^2))$ over K; then we must also have $k((G^3)) \subset K((q^1), (q^2))$. Since both $K((p^1), (p^2))$ and $K((q^1), (q^2))$ are regular extensions of K and since they are independent with respect to each other over K, it follows that $K((p^1), (p^2)) \cap K((q^1), (q^2)) = K$. Therefore we conclude that $k((G^3)) \subset K = k((G^1), (G^2))$. In order to prove that the addition on $\mathfrak{V}$ is a rational function, we have only to show that $(G^a) + (G^b)$ is the only specialization of $(G^1) + (G^2)$ over any specialization $(G^1) \to (G^a)$, $(G^2) \to (G^b)$ over k. Here again our assertion will follow immediately from the fact that the relation $(G^c) = (G^a) + (G^b)$ is an algebraic condition over k between the three points. To show this, let $\mathfrak{P}^a$, $\mathfrak{P}^b$, $\mathfrak{P}^c$ be the divisor classes determined by the points (G^a), (G^b), (G^c) respectively, and let $\mathfrak{p}^a$, $\mathfrak{p}^b$, $\mathfrak{p}^c$ be any positive divisors in the classes $\mathfrak{P}^a$, $\mathfrak{P}^b$, $\mathfrak{P}^c$ respectively; let $\mathfrak{o}$ be a rational positive divisor over k in the divisor class $\mathfrak{D}$. We recall again ([7], § 14) that linear equivalence between positive divisors is an algebraic condition over k; since the positive divisor $\mathfrak{o}$ is rational over k and hence is represented by a rational point over k in $\mathfrak{C}^n$, the relation $\mathfrak{p}^a + \mathfrak{p}^b \sim \mathfrak{p}^c + \mathfrak{o}$ can be expressed as an algebraic condition over k in terms of the points (p^a), (p^b), (p^c). It follows then by Lemma 1 (the case where the points are not involved at all) that the relation $(G^c) = (G^a) + (G^b)$ is an algebraic condition over k.

Thus we have shown that the addition in $\mathfrak{D}_n$ (or the addition in $\mathfrak{D}_0$) induces on the variety $\mathfrak{V}$ an addition of points, which is a rational function, defined over the ground field k, on the product variety $\mathfrak{V} \times \mathfrak{V}$ with values in $\mathfrak{V}$. Moreover, this function is not only defined for a generic point of $\mathfrak{V} \times \mathfrak{V}$, but also has a uniquely determined value for any point on $\mathfrak{V} \times \mathfrak{V}$. From this it follows that the points $\mathfrak{V}$ form an abelian group under this addition operation, with the point (G^0) as the unit element. The inverse of any point (G) corresponding to the class $\mathfrak{P}$ is the point $(G)^{-1}$ corresponding to the class $2\mathfrak{D} - \mathfrak{P}$; and in case (G) is a generic point of V over k, the point $(G)^{-1}$ is rational over $k((G))$. From this it follows, by a similar

argument as before, that the inverse of a point is a rational function on $\mathfrak{V}$, defined over k. We shall write $(G^1) - (G^2)$ for $(G^1) + (G^2)^{-1}$. Finally, it is also easily seen that, for any fixed point (G^a) on $\mathfrak{V}$, the mapping $(G) \to (G^a) + (G)$ defines a birational transformation of $\mathfrak{V}$ onto itself, which is one-to-one for every point of $\mathfrak{V}$. In order to prove that $\mathfrak{V}$ is an Abelian variety, it remains to shows that the function $(G^1) + (G^2)$ is defined at every point of the product variety $\mathfrak{V} \times \mathfrak{V}$; since a rational function is defined at every simple point for which it has a uniquely determined value, it is sufficient to show that the variety $\mathfrak{V}$ is non-singular. This will be done in section 5, where this result is obtained as an application of a general theorem proved by us elsewhere; in the final section we shall then introduce the canonical homomorphism and show that it has the desired properties, thus completing the proof that $\mathfrak{V}$ is a Jacobian variety of the curve $\mathfrak{C}$. Before we proceed with this, we shall insert a digression in the next section, where we shall show that a derived (absolutely) normal model $\mathfrak{W}$ of the variety $\mathfrak{V}$ is an Abelian variety and that the birational correspondence between $\mathfrak{V}$ and $\mathfrak{W}$ is one-to-one without exception; this result, together with the results of the final section on the canonical homomorphism, gives us an alternate proof of the existence of a Jacobian variety, though not necessarily defined over the original ground field k, except in case k is perfect. Although this alternate proof is now superseded by the more complete result in section 5, it still has the advantage of simplicity; furthermore, the method of proof used here may have possible applications to other similar problems.*

4. The derived normal model $\mathfrak{W}$. Consider a derived (absolutely) normal variety $\mathfrak{W}$ of $\mathfrak{V}$, which is defined over a purely inseparable extension K of the field k. The birational correspondence $\Gamma(\mathfrak{W}) = \mathfrak{V}$ between $\mathfrak{W}$ and $\mathfrak{V}$, also defined over K, has the property that to each point (H) of $\mathfrak{W}$ corresponds exactly one point $\Gamma(H)$ of $\mathfrak{V}$, while conversely to each point (G) of $\mathfrak{V}$ correspond at most a finite number of points $\Gamma^{-1}(G)$ of $\mathfrak{W}$. Let (H^1) and (H^2) be two independent generic points on $\mathfrak{W}$ over K, and let $(G^1) = \Gamma(H^1)$

* [*Added in proof, March* 10, 1954] T. Matsusaka, in a recent paper "On the algebraic construction of Picard variety (II)," *Japanese Journal of Mathematics*, vol. 22, pp. 51-62, has successfully applied this method to prove the existence of the Picard variety over an arbitrary field; the application of this method was made possible by a crucial result, proved by Matsusaka in another paper, concerning the existence of a "regular maximal algebraic family" of divisors. Matsusaka has also shown that it is sufficient to take a derived normal model with reference to k, which can then be shown to be also absolutely normal; this removes the restriction on the field of definition stated above.

and $(G^2) = \Gamma(H^2)$ be the corresponding points on $\mathfrak{V}$ under the correspondence Γ; then (G^1) and (G^2) are also generic and independent with respect to each other over K. The sum $(G^3) = (G^1) + (G^2)$ of the two independent generic points (G^1) and (G^2) on $\mathfrak{V}$, being a rational function on $\mathfrak{V} \times \mathfrak{V}$ defined over K, is carried by the birational coorespondence Γ into the rational function $(H^3) = \Gamma^{-1}(G^3) = \Gamma^{-1}(\Gamma(H^1) + \Gamma(H^2))$ on $\mathfrak{W} \times \mathfrak{W}$, also defined over K, which we shall define as the sum $(H^1) + (H^2)$ of the two independent generic points (H^1) and (H^2) on $\mathfrak{W}$. Now, what is the behavior of this function $(H^1) + (H^2)$ for a pair of special points (H^a) and (H^b) on $\mathfrak{W}$? By definition, any value of the function $(H^1) + (H^2)$ is a specialization (H^c) of the point (H^3) over the specialization $(H^1) \to (H^a)$, $(H^2) \to (H^b)$ over K. It is well known that any such specialization $(H^1, H^2, H^3) \to (H^a, H^b, H^c)$ can always be extended to a specialization of (G^1, G^2, G^3), while on the other hand it is clear from the above that any such extension of specialization is already uniquely determined by the specialization $(H^1, H^2) \to (H^a, H^b)$ alone. In fact, since the transformation $\Gamma(H)$ has always uniquely determined specialization at every point of $\mathfrak{W}$, we have $(G^1) \to \Gamma(H^a)$, $(G^2) \to \Gamma(H^b)$; from this it follows that since the sum $(G^1) + (G^2)$ has always a uniquely determined specialization for every pair of points on $\mathfrak{V}$, we must have $(G^3) \to \Gamma(H^a) + \Gamma(H^b)$. Since, in particular, the relation $\Gamma(H^3) = (G^3)$, which defines the correspondence Γ, must be preserved under the extended specialization, we have the relation $\Gamma(H^c) = \Gamma(H^a) + \Gamma(H^b)$. It follows then that any specialization (H^c) of (H^3), over the specialization $(H^1, H^2) \to (H^a, H^b)$ over K, must be one of the finite number of points of $\mathfrak{W}$ which are carried over by Γ into the point $\Gamma(H^a) + \Gamma(H^b)$ on $\mathfrak{V}$. In similar manner, we can consider a specialization (H^b) of the point $(H^2) = \Gamma^{-1}(\Gamma(H^3) - \Gamma(H^1))$ over any given specialization $(H^1, H^3) \to (H^a, H^c)$, where the (H^1) and (H^3) are now a pair of independent generic points of $\mathfrak{W}$. Since the specialization can also be extended in a unique way to the specialization $(G^1, G^3, G^2) \to (\Gamma(H^a), \Gamma(H^c), \Gamma(H^c) - \Gamma(H^a))$, we conclude that the relation $\Gamma(H^b) = \Gamma(H^c) - \Gamma(H^a)$ must hold for any such specialization, and hence (H^b) must be one of the finite number of points of $\mathfrak{W}$ which are carried over by Γ into the point $\Gamma(H^c) - \Gamma(H^a)$ of $\mathfrak{V}$.

Thus we have shown that for any specialization $(H^1, H^2, H^3) \to (H^a, H^b, H^c)$ the relation $\Gamma(H^c) = \Gamma(H^a) + \Gamma(H^b)$ always holds, and that in this specialization any two of the three points can be chosen arbitrarily on $\mathfrak{W}$, while the third one is then one of the finite number of possible points determined by this relation. Let now (H^a) be any given point of $\mathfrak{W}$, and let (H) be a

generic point of $\mathfrak{W}$ over $K((H^a))$. Since the point $\Gamma(H^a) + \Gamma(H)$ is a generic point of $\mathfrak{V}$ over $K((H^a))$, the relation $\Gamma(H') = \Gamma(H^a) + \Gamma(H)$ determines uniquely a specialization $(H^3) \to (H')$ over the specialization $(H^1, H^2) \to (H^a, H)$ over K, and this point (H') is also a generic point of $\mathfrak{W}$ over $K((H^a))$. Conversely, given a generic point (H') of $\mathfrak{W}$ over $K((H^a))$, the same relation determines uniquely a specialization $(H^2) \to (H)$ over the specialization $(H^1, H^3) \to (H^a, H')$ over K, and the point (H) is also generic over $K((H^a))$. In other words, for any given point (H^a) of $\mathfrak{W}$, the relation $(H^a) + (H) = (H')$ determines a birational correspondence $\Theta_a(H) = (H')$ of the variety $\mathfrak{W}$ onto itself; in fact we have $(H') = \Gamma^{-1}(\Gamma(H^a) + \Gamma(H))$ for any generic (H) over $K((H^a))$ and $(H) = \Gamma^{-1}(\Gamma(H') - \Gamma(H^a))$ for generic (H') over $K((H^a))$. This birational correspondence Θ_a between the normal variety $\mathfrak{W}$ and itself has the property that to each specialization of (H) correspond at most a finite number of specializations of (H'), and vice versa. For, any specialization $(H, H') \to (H^b, H^c)$ of this correspondence Θ_a over K is contained in a specialization $(H^1, H^2, H^3) \to (H^a, H^b, H^c)$ over K, and we have seen above that there is only a finite number of the latter when two of the three points (H^a, H^b, H^c) are given. This means that the birational correspondence Θ_a has no fundamental points on $\mathfrak{W}$, neither for the direct transformation Θ_a nor for the inverse transformation Θ_a^{-1}. Since for a birational transformation, or more generally for a rational transformation of a normal variety, a point is either fundamental or regular ([10], Th. 9 and Th. 10, note the difference in terminology), we conclude that the birational correspondonce Θ_a must be biregular everywhere on $\mathfrak{W}$.

From the biregularity of the correspondence Θ_a for every point (H^a) we can draw some important conclusions. Let (H^c) be a simple point on $\mathfrak{W}$ with the property that it is the only point on $\mathfrak{W}$ which is carried by Γ into the point $\Gamma(H^c)$ on $\mathfrak{V}$; we can take, for example, a generic point of $\mathfrak{W}$ over K. Given any point (H^b), let (H^a) be any point such that $\Gamma(H^c) = \Gamma(H^a) + \Gamma(H^b)$. The birational correspondence Θ_a, determined by the so chosen point (H^a), will then carry the point (H^b) into a point with the image $\Gamma(H^a) + \Gamma(H^b)$ on $\mathfrak{V}$ under the correspondence Γ, which can only be the point (H^c). Since the correspondence Θ_a is biregular and the point (H^c) is simple, it follows that the point (H^b) is also simple. As (H^b) is any point on $\mathfrak{W}$, this shows that the variety $\mathfrak{W}$ is free of singularities. Furthermore, let (H^β) be another point such that $\Gamma(H^\beta) = \Gamma(H^b)$, then the same birational correspondence Θ_a must also carry (H^β), into (H^c),

from which it follows, on account of the biregularity of Θ_a, that $(H^\beta) = (H^\nu)$. This shows that the correspondence Γ between $\mathfrak{W}$ and $\mathfrak{V}$ is one-to-one for all points without exception.

Thus we have shown the existence of an Abelian variety $\mathfrak{W}$ which is in one-to-one birational correspondence without exception with the variety $\mathfrak{V}$; as we have mentioned in our paper [1], this is our original proof of the existence of a Jacobian variety for the curve $\mathfrak{C}$. The underlying idea of the proof is the simple observation that if a variety is homogeneous in the sense that there is a transitive group of rational transformations which are everywhere determined, and if the variety is normal, then it is non-singular. That a variety which is homogeneous in this sense, but not normal, need not be non-singular, is shown by the simple example of a plane cubic curve with a cusp, which is in one-to-one birational correspondence without exception with the projective line. The question naturally arises as to whether the variety $\mathfrak{V}$ is itself also non-singular and is already a Jacobian variety. This question, besides being of some interest in itself, is significant not only on account of the fact that $\mathfrak{V}$ is defined over k, but also on account of the fact, proved in the next section, that for any point (G) in $\mathfrak{V}$ the corresponding variety $\mathfrak{G}$ in $\mathfrak{C}^n$ is defined over the field $k((G))$.

5. Proof that the variety $\mathfrak{V}$ is non-singular. Consider the rational transformation ϕ of $\mathfrak{C}^n$ onto $\mathfrak{V}$, defined over k; if (z) is a generic point of $\mathfrak{V}$ over k, then the inverse image $\phi^{-1}(z)$ is a subvariety $\mathfrak{G}(z)$ of dimension $n-g$ in $\mathfrak{C}^n$, defined over $k((z))$, and for any specialization $(z) \to (\zeta)$ over k, the variety $\mathfrak{G}(z)$, considered as a positive cycle of dimension $n-g$, has a uniquely determined specialization $\mathfrak{G}(\zeta)$, which is a multiple of the variety $\phi^{-1}(\zeta)$. In the terminology introduced in our paper [1], the rational transformation ϕ induces an involutional system of positive cycles $|\mathfrak{G}(z)|$ on $\mathfrak{C}^n$, and the variety $\mathfrak{V}$ is the associated variety of this involutional system; according to the main theorem in that paper ([1], p. 258), the variety $\mathfrak{V}$ is simple at every point (ζ) for which the positive cycle $\mathfrak{G}(\zeta)$ coincides with the variety $\phi^{-1}(\zeta)$ and contains a simple point in $\mathfrak{C}^n$. In order to prove that the variety $\mathfrak{V}$ is non-singular, it is therefore sufficient to show that (1) the variety $\phi^{-1}(\zeta)$ has the same degree (as a subvariety in $\mathfrak{S}_t$) for every point (ζ) in $\mathfrak{V}$, i. e. the degree of the variety $\mathfrak{G}_\mathfrak{P}$ is the same for every divisor class $\mathfrak{P}$ of $\mathfrak{C}$, and (2) the variety $\mathfrak{C}^n$ is non-singular.

We shall show that the degree of the variety $\mathfrak{G}_\mathfrak{P}$, for any divisor class $\mathfrak{P}$ of degree n, is equal to h^{n-g}. By definition, the degree of $\mathfrak{G}_\mathfrak{P}$ is the

number of intersections of $\mathfrak{G}_{\mathfrak{P}}$ with $n-g$ independent generic hyperplanes $\sum_{j=0}^{t} w^i{}_j Y_j = 0$ $(i = 1, \cdots, n-g)$ in $\mathfrak{S}_t$ over K, where K is a field over which the divisor class $\mathfrak{P}$ is rational. Consider the $n-g$ independent generic hyperplanes $\sum_{j=0}^{m} v^i{}_j X_j = 0$ $(i = 1, \cdots, n-g)$ in $\mathfrak{S}_m$ over K, and for each i $i = 1, \cdots, n-g$, let $(\xi^{i,1}), \cdots, (\xi^{i,h})$ be the h intersections of the i-th hyperplane with the curve $\mathfrak{C}$; let $\mathfrak{p}^\lambda$, where $\lambda = (m_1, \cdots, m_{n-g})$ runs through all the h^{n-g} possible ordered sets of $n-g$ positive integers $m_i \leq h$, be the positive divisor of the class $\mathfrak{P}$ which contains the $n-g$ points (ξ^{i,m_i}), $i = 1, \cdots, n-g$. It is then easily seen that the h^{n-g} points (p^λ) in $\mathfrak{G}_{\mathfrak{P}}$ corresponding to the h^{n-g} positive divisors $\mathfrak{p}^\lambda$ constitute all the intersections of $\mathfrak{G}_{\mathfrak{P}}$ with the $n-g$ special hyperplanes $\sum_{j=0}^{t} \Omega_j(v^i) Y_j = 0$ $(i = 1, \cdots, n-g)$ in $\mathfrak{S}_t$. In fact, let (q) be the point in $\mathfrak{C}^n$ corresponding to a positive divisor $\mathfrak{q}$ of degree n consisting of the n points $(\eta^1), \cdots, (\eta^n)$, then we have $\sum_{j=0}^{t} \Omega_j(v^i) q_j = \prod_{k=1}^{n} \left(\sum_{j=0}^{m} v^i{}_j \eta^k{}_j \right)$, $i = 1, \cdots, n-g$. Hence, all the h^{n-g} points (p^λ) of $\mathfrak{G}_{\mathfrak{P}}$ lie on the $n-g$ special hyperplanes; and conversely, if a point of $\mathfrak{G}_{\mathfrak{P}}$ lies on these $n-g$ special hyperplanes, then the corresponding positive divisors must contain at least one point from each of the $n-g$ sets $(\xi^{i,1}), \cdots, (\xi^{i,h})$, $i = 1, \cdots, n-g$, and hence must be one of the h^{n-g} positive divisors $\mathfrak{p}^\lambda$. Therefore, in order to prove our assertion, we need only show that each point (p^λ) is a simple intersection of $\mathfrak{G}_{\mathfrak{P}}$ with these $n-g$ special hyperplanes. Since the point (p^λ) is a generic point of $\mathfrak{G}_{\mathfrak{P}}$ over K (it has the dimension $n-g$ over K), it is a simple point of $\mathfrak{G}_{\mathfrak{P}}$; hence, we need only show that the linear variety of dimension $t-n+g$ in $\mathfrak{S}_t$ defined by the $n-g$ special hyperplanes is transversal to the tangent linear variety of $\mathfrak{G}_{\mathfrak{P}}$ at the point (p^λ). As we shall from now on only deal with the one divisor $\mathfrak{p}^\lambda$ or point (p^λ), for an arbitrary but fixed λ, we can drop the superscript λ and simply write $\mathfrak{p}$ or (p).

Let $(x^1), \cdots, (x^n)$ be the n points contained in the divisor $\mathfrak{p}$, where we can assume that the first $n-g$ points are the points (ξ^{i,m_i}), $i = 1, \cdots, n-g$. Since $\mathfrak{p}$ is a generic positive divisor of the class $\mathfrak{P}$ over K, all the n points $(x^1), \cdots, (x^n)$ are generic points of the curve $\mathfrak{C}$ over K and distinct from each other; moreover, any $n-g$ of these points are independent with respect to each other over K and determine the remaining g points uniquely. Hence, there is an automorphism of the field $K((x^1), \cdots, (x^n))$ over the field $K((p))$ which carries any set of $n-g$ of the n points $(x^1), \cdots, (x^n)$ into

any other such set. If $\sum_{j=0}^{m} v_j X_j = 0$ is a hyperplane in $\mathfrak{S}_m$ containing one of the n points $(x^1), \cdots, (x^n)$, then the hyperplane $\sum_{j=0}^{t} \Omega_j(v) Y_j = 0$ in $\mathfrak{S}_t$ evidently contains the point (p); we shall call such a hyperplane in $\mathfrak{S}_t$ a derived hyperplane of the hyperplane in $\mathfrak{S}_m$. The $n-g$ special hyperplanes $\sum_{j=0}^{t} \Omega_j(v^i) Y_j = 0$ $(i = 1, \cdots, n-g)$ considered above are all derived hyperplanes, since each hyperplane $\sum_{j=0}^{m} v^i{}_j X_j = 0$ in $\mathfrak{S}_m$ contains the point (x^i), $i = 1, \cdots, n-g$. Moreover, for each $i = 1, \cdots, n-g$, the hyperplane $\sum_{j=0}^{m} v^i{}_j X_j = 0$ is a generic hyperplane over $K((x^1), \cdots, (x^n))$ of the system of all hyperplanes in $\mathfrak{S}_m$ which contain the point (x^i); and the $n-g$ hyperplanes $\sum_{j=0}^{m} v^i{}_j X_j = 0$ $(i = 1, \cdots, n-g)$, considered as generic hyperplanes of the $n-g$ corresponding systems over $K((x^1), \cdots, (x^n))$, are independent with respect to each other over $K((x^1), \cdots, (x^n))$.

Consider now the set of all derived hyperplanes in $\mathfrak{S}_t$; it can be easily seen that the point (p) is the only point in $\mathfrak{S}_t$ common to all these hyperplanes. For, each point in $\mathfrak{S}_t$ corresponds to a form of degree n in (V), and the form $\sum_{j=0}^{t} p_j \Omega_j(V) = \prod_{i=1}^{n} (\sum_{j=0}^{m} x^i{}_j V_j)$ is obviously the only form of degree n in (V) which vanishe for all the solutions of any one of the n linear equations $\sum_{j=0}^{m} x^i{}_j V_j = 0$, $i = 1, \cdots, n$. Thus there exist t linearly independent derived hyperplanes in $\mathfrak{S}_t$, and consequently there is a set of $n-g$ hyperplanes $\sum_{j=0}^{m} v^i{}_j X_j = 0$ $(i = 1, \cdots, n-g)$ in $\mathfrak{S}_m$, each containing a point (x^{l_i}) of the divisor $\mathfrak{p}$, such that the $n-g$ derived hyperplanes $\sum_{j=0}^{t} \Omega_j(v^i) Y_j = 0$ $(i = 1, \cdots, n-g)$ define a linear variety of dimension $t-n+g$ transversal to the tangent linear variety of $\mathfrak{G}_{\mathfrak{P}}$ at (p). The $n-g$ points (x^{l_i}), $i = 1, \cdots, n-g$, are distinct; for, otherwise there would be infinitely many positive divisors in the class $\mathfrak{P}$ which contain all the (less than $n-g$ distinct) points (x^{l_i}), $i = 1, \cdots, n-g$, and the corresponding points in $\mathfrak{G}_{\mathfrak{P}}$ would then be contained in the $n-g$ derived hyperplanes $\sum_{j=0}^{t} \Omega_j(v^i) Y_j = 0$ $(i = 1, \cdots, n-g)$, and consequently these derived hyperplanes could not define a linear variety transversal to the tangent linear variety of $\mathfrak{G}_{\mathfrak{P}}$ at (p). Thus, for each $i = 1, \cdots, n-g$, the set

(v^i) is a solution of the linear equation $\sum_{j=0}^{m} x^{l_i}{}_j V_j = 0$. Let $(\bar{v}^i)$, $i = 1, \cdots, n-g$, be respectively generic solutions of the linear equations $\sum_{j=0}^{m} x^{l_i}{}_j V_j = 0$, $i = 1, \cdots, n-g$, over $K((x^1), \cdots, (x^n))$, independent with respect to each other over $K((x^1), \cdots, (x^n))$; then the $n-g$ derived hyperplanes $\sum_{j=0}^{t} \Omega_j(\bar{v}^i) Y_j = 0$ $(i = 1, \cdots, n-g)$ define a linear variety which is also transversal to the tangent linear variety of $\mathfrak{G}_{\mathfrak{P}}$ at (p). For, the set of $n-g$ hyperplanes $\sum_{j=0}^{t} \Omega_j(v^i) Y_j = 0$ $(i = 1, \cdots, n-g)$ is a specialization over $K((x^1), \cdots, (x))$ of the set of $n-g$ hyperplanes $\sum_{j=0}^{t} \Omega_j(\bar{v}^i) Y_j = 0$ $(i = 1, \cdots, n-g)$, and the former set defines a linear variety transversal to the tangent linear variety of $\mathfrak{G}_{\mathfrak{P}}$ at (p). Now, there is an automorphism of the field $K((x^1), \cdots, (x^n))$ over $K((p))$, which carries the $n-g$ points $(x^{l_1}), \cdots, (x^{l_{n-g}})$ into the $n-g$ points $(x^1), \cdots, (x^{n-g})$ respectively, and this automorphism can be extended into an isomorphism between the fields $K((x^1), \cdots, (x^n), (\bar{v}^1), \cdots, (\bar{v}^{n-g}))$ and $K((x^1), \cdots, (x^n), (v^1), \cdots, (v^{n-g}))$ by the correspondence $(\bar{v}^i) \leftrightarrow (v^i)$, $i = 1, \cdots, n-g$. In this isomorphism the derived hyperplanes $\sum_{j=0}^{t} \Omega_j(\bar{v}^i) Y_j = 0$ $(i = 1, \cdots, n-g)$ are carried over into the derived hyperplanes $\sum_{j=0}^{t} \Omega_j(v^i) Y_j = 0$ $(i = 1, \cdots, n-g)$; since the former set defines a linear variety of $\mathfrak{G}_{\mathfrak{P}}$ at (p), the same must be true of the latter. This completes the proof of the assertion that the degree of $\mathfrak{G}_{\mathfrak{P}}$ is h^{n-g}.

Next, we shall show that the variety $\mathfrak{C}^n$ is non-singular. We observe first that in a similar manner as above, by taking (p) to be a generic point of $\mathfrak{C}^n$ over k and considering n instead of $n-g$ derived hyperplanes in $\mathfrak{S}_t$, we can show that the degree of the variety $\mathfrak{C}^n$ is equal to h^n. In fact, the proof in this case is somewhat simpler than the above, for there it is not necessary to consider any automorphisms of the field $k((x^1), \cdots, (x^n))$. Now, let (q) be any point in $\mathfrak{C}^n$ and let the corresponding positive divisor $\mathfrak{q}$ be composed of the n (not necessarily distinct) points $(\eta^1), \cdots, (\eta^n)$ in $\mathfrak{C}$. Consider the n linear equations $\sum_{j=0}^{m} \eta^i{}_j V_j = 0$ $(i = 1, \cdots, n)$; let (ρ^i) $(i = 1, \cdots, n)$ be n generic solutions of these n equations, respectively, over $k((\eta^1), \cdots, (\eta^n))$, independent with respect to each other over $k((\eta^1), \cdots, (\eta^n))$; for each $i = 1, \cdots, n$, let $(\eta^{i,1}) = (\eta^i), (\eta^{i,2}), \cdots, (\eta^{i,h})$

be the h intersections of the hyperplane $\sum_{j=0}^{m} \rho^i{}_j X_j = 0$ with the curve $\mathfrak{C}$. Since $\mathfrak{C}$ is non-singular (and the hyperplanes have been chosen as above), the points $(\eta^{i,j})$ $(i = 1, \cdots, n; j = 2, \cdots, h)$ are all distinct from each other and different from the points $(\eta^{i,1})$ $(i = 1, \cdots, n)$. It follows then that all the h^n positive divisors $\mathfrak{q}^\lambda = ((\eta^{1,m_1}), \cdots, (\eta^{n,m_n}))$, where the $\lambda = (m_1, \cdots, m_n)$ runs through all the h^n possible ordered sets of n positive integers $m_i \leq h$ $(i = 1, \cdots, n)$, are distinct from each other. It is easily seen that the h^n corresponding distinct points (q^λ) are precisely the intersections of $\mathfrak{C}^n$ with the linear variety of dimension $t - n$ in $\mathfrak{S}_t$ defined by the hyperplanes $\sum_{j=0}^{t} \Omega_j(\rho^i) Y_j = 0$ $(i = 1, \cdots, n)$. Since the degree of $\mathfrak{C}^n$ is h^n, it follows that each point (q^λ) and hence in particular the point (q) is a simple intersection of $\mathfrak{C}^n$ with this linear variety. This shows that (q) is a simple point in $\mathfrak{C}^n$, and as (q) is an arbitrary point in $\mathfrak{C}^n$, this means that the variety $\mathfrak{C}^n$ is non-singular.

Finally, we observe that in exactly the same manner as above, we can show that there exists a linear variety of dimension $t - n + g$ in $\mathfrak{S}_t$ which intersects simply any given $\mathfrak{G}_\mathfrak{B}$ at any given point (q); it follows then that the variety $\mathfrak{G}_\mathfrak{B}$ is also non-singular for every $\mathfrak{B}$.

6. The canonical homomorphism. Consider again the rational transformation ϕ of $\mathfrak{C}^n$ onto $\mathfrak{V}$; we shall show that it generates in a natural way (by linear extension, in the terminology of Weil) a rational homomorphism Φ of $\mathfrak{G}_a(\mathfrak{C})$ onto the Abelian variety $\mathfrak{V}$ with the subgroup $\mathfrak{G}_l(\mathfrak{C})$ as the kernel, and that the so defined homomorphism Φ is defined over k and has the "universal mapping" property. We observe first that if $\mathfrak{p}^1, \cdots, \mathfrak{p}^s$ and $\mathfrak{q}^1, \cdots, \mathfrak{q}^s$ are two sets of positive divisors of degree n in $\mathfrak{C}$, then we have the relation $\sum_{i=1}^{s} \phi(p^i) = \sum_{i=1}^{s} \phi(q^i)$ if and only if $\sum_{i=1}^{s} \mathfrak{p}^i \sim \sum_{i=1}^{s} \mathfrak{q}^i$; in fact, let $\mathfrak{o}$ be any divisor in the class $\mathfrak{O}$ and let $\mathfrak{p}^0$ and $\mathfrak{q}^0$ be respectively positive divisors such that the relations $\sum_{i=1}^{s} \mathfrak{p}^i \sim (s-1)\mathfrak{o} + \mathfrak{p}^0$ and $\sum_{i=1}^{s} \mathfrak{q}^i \sim (s-1)\mathfrak{o} + \mathfrak{q}^0$ hold respectively, then we have, by definition, the relations $\phi(p^0) = \sum_{i=1}^{s} \phi(p^i)$ and $\phi(q^0) = \sum_{i=1}^{s} \phi(q^i)$, and it is clear that $\phi(p^0) = \phi(q^0)$ if and only if $\sum_{i=1}^{s} \mathfrak{p}^i \sim \sum_{i=1}^{s} \mathfrak{q}^i$. We shall call this relation the compatibility condition for

the transformation ϕ. Now, let $(\xi^1), \cdots, (\xi^{n-1})$ be $n-1$ independent generic points over k in $\mathfrak{C}$, and let (x) be a generic point of $\mathfrak{C}$ over $k((\xi^1), \cdots, (\xi^{n-1}))$; let (y) be the point in $\mathfrak{C}^n$ representing the divisor consisting of the points $(\xi^1), \cdots, (\xi^{n-1})$ and (x), and let ϕ' be the rational transformation of $\mathfrak{C}$ into $\mathfrak{V}$ determined by the correspondence $(x) \to \phi(y)$, which is defined over $k((\xi^1), \cdots, (\xi^{n-1}))$. For any divisor $\mathfrak{p} = \sum_{i=1}^{s} n_i(x^i)$ in $\mathfrak{G}_a(\mathfrak{C})$, we define $\Phi(\mathfrak{p}) = \sum_{i=1}^{s} n_i \phi'(x^i)$; it is clear that the so defined mapping Φ is a rational homomorphism of $\mathfrak{G}_a(\mathfrak{C})$ into $\mathfrak{V}$, defined over $k((\xi^1), \cdots, (\xi^{n-1}))$. It can be easily seen from the compatibility condition that the kernel of the homomorphism Φ is precisely the subgroup $\mathfrak{G}_1(\mathfrak{C})$, and that this definition of Φ is actually independent of the choice of the points $(\xi^1), \cdots, (\xi^{n-1})$. If $(\eta^1), \cdots, (\eta^{n-1})$ are $n-1$ independent generic points of $\mathfrak{C}$ over k, independent with respect to the points $(\xi^1), \cdots, (\xi^{n-1})$ over k, then the rational homomorphism Φ is also defined over $k((\eta^1), \cdots, (\eta^{n-1}))$; it follows then that Φ must be defined over

$$k((\xi^1), \cdots, (\xi^{n-1})) \cap k((\eta^1), \cdots, (\eta^{n-1})) = k.$$

In fact, if $\mathfrak{p}$ is any divisor in $\mathfrak{C}$, rational over any extension K of k, then the point $\Phi(\mathfrak{p})$ is rational over both $K((\xi^1), \cdots, (\xi^{n-1}))$ and $K((\eta^1), \cdots, (\eta^{n-1}))$ and hence must be rational over

$$K((\xi^1), \cdots, (\xi^{n-1})) \cap K((\eta^1), \cdots, (\eta^{n-1})) = K.$$

Consider now a rational homomorphism Ψ of $\mathfrak{G}_a(\mathfrak{C})$ into an Abelian variety $\mathfrak{A}$, defined over an extension K of k; let (z) be a generic point of $\mathfrak{V}$ over K, and let (q) be a generic point of the variety $\phi^{-1}(z)$ over $K((z))$ and $\mathfrak{q}$ be the positive divisor represented by (q). Since the divisor class $\mathfrak{O}$ is rational over K, it contains a divisor $\mathfrak{o}$ which is rational over K; the point $\Psi(\mathfrak{q} - \mathfrak{o})$ is then rational over $K((q), (z))$. The correspondence $(q) \to \Psi(\mathfrak{q} - \mathfrak{o})$ then determines over $K((z))$ a rational transformation of $\phi^{-1}(z)$ into $\mathfrak{A}$; since $\phi^{-1}(z)$ is birationally equivalent to a projective space, it follows from a result of Weil ([9], Th. 8, Cor.) that this rational transformation must be a constant, so that the point $\Psi(\mathfrak{q} - \mathfrak{o})$ is rational over $K((z))$. The correspondence $(z) \to \Psi(\mathfrak{q} - \mathfrak{o})$ then determines a rational transformation Ψ_1 of $\mathfrak{V}$ into $\mathfrak{A}$, defined over K, which carries the unit element in $\mathfrak{V}$ into the unit element in $\mathfrak{A}$; according to another result of Weil ([9], Th. 9), this rational transformation Ψ_1 must be a homomorphism of $\mathfrak{V}$ into $\mathfrak{A}$. Let $\mathfrak{p} = \sum_{i=1}^{s} n_i(x^i)$ be again any divisor in $\mathfrak{G}_a(\mathfrak{C})$, and for each

$i = 1, \cdots, n$, let $\mathfrak{y}^i$ be the positive divisor of degree n consisting of the points $(\xi^1), \cdots, (\xi^{n-1})$ and (x^i); since $\sum_{i=1}^{s} n_i = 0$, we have evidently the equation $\mathfrak{p} = \sum_{i=1}^{s} n_i(\mathfrak{y}^i - \mathfrak{o})$. We have then

$$\Psi(\mathfrak{p}) = \sum_{i=1}^{s} n_i \Psi(\mathfrak{y}^i - \mathfrak{o}) = \sum_{i=1}^{s} n_i \Psi_1(\phi'(x^i)) = \Psi_1\Big(\sum_{i=1}^{s} n_i \phi'(x^i)\Big) = \Psi_1\Phi(\mathfrak{p}).$$

Thus we have shown that Φ is the canonical homomorphism, and this completes the proof that $\mathfrak{V}$ is the Jacobian variety of the curve $\mathfrak{C}$.

Once we have obtained the canonical homomorphism Φ, we can derive the rational transformation ϕ from it by the formula $\phi(q) = \Phi(\mathfrak{q} - \mathfrak{o})$, and this formula holds for every positive integer $n > 2g - 2$, provided there exists a rational divisor $\mathfrak{o}$ over k of degree n. In fact, in order that the rational transformation ϕ be defined over k, it is sufficient that $\mathfrak{o}$ is contained in a class $\mathfrak{O}$ such that the variety $\mathfrak{G}_{\mathfrak{O}}$ is defined over k. In general, we shall say that a divisor class $\mathfrak{P}$ in $\mathfrak{D}_n$ $(n > 2g - 2)$ is rationally defined over a field K if the variety $\mathfrak{G}_{\mathfrak{P}}$ is defined over K; and an arbitrary divisor class is said to be rationally defined over K if it can be expressed as the difference of two divisor classes which are rationally defined over K. It is clear that a rational divisor class over K is also rationally defined over K, but the converse of this statement is not generally true, as can be shown by examples. It can be easily shown that, for any positive integer n, if there exists a rationally defined divisor class $\mathfrak{O}$ over k of degree n, then the formula $\phi(q) = \Phi(\mathfrak{q} - \mathfrak{o})$ determines a rational transformation ϕ of $\mathfrak{C}^n$ into $\mathfrak{V}$, defined over k. We maintain now that, conversely, if there is a rational transformation of $\mathfrak{C}^n$ into $\mathfrak{V}$, defined over k, which generates by linear extension the canonical homomorphism Φ, then there exists a rationally defined divisor class over k of degree n in $\mathfrak{C}$. In fact, for every positive integer s, the rational transformation ϕ induces a rational transformation ϕ_s of $\mathfrak{C}^{ns}$ into $\mathfrak{V}$, determined by the formula $\phi_s(p) = \sum_{i=1}^{s} \phi(p^i)$, where (p) is any point in $\mathfrak{C}^{ns}$ and $(p^1), \cdots, (p^s)$ are s points in $\mathfrak{C}^n$ such that we have $\mathfrak{p} = \sum_{i=1}^{s} \mathfrak{p}^i$ for the corresponding positive divisors, the particular choice of the points $(p^1), \cdots, (p^s)$ being immaterial on account of the compatibility condition. It is clear that the rational transformation ϕ_s is defined over k, and that, again by virtue of the compatibility condition, we have the equation $\phi_s^{-1}(\phi'(p)) = \mathfrak{G}_{\mathfrak{P}}$, where $\mathfrak{P}$ is the divisor class determined by $\mathfrak{p}$. It is also clear that for sufficiently large s, say $s \geq g$, the rational transformation ϕ_s is onto; if $\mathfrak{P}_s$ is the divisor class

of degree ns such that $\phi_s^{-1}(0) = \mathfrak{G}_{\mathfrak{P}_s}$, then $\mathfrak{P}_s$ is evidently rationally defined over k. The divisor class $\mathfrak{P}_{s+1} - \mathfrak{P}_s$, for any sufficiently large s, is then a rationally defined divisor class over k of degree n. In case $n = 1$, we have the "canonical function" of Weil, and our result offers an answer to the question raised by Weil ([9], p. 68) as to when a field of definition is "complete" for a curve.

Finally, we shall add a remark concerning the special case when the curve $\mathfrak{C}$ contains a rational divisor over k of degree 1; such is always the case, for example, when the field k is finite, as has been shown by F. K. Schmidt ([6], p. 27). We shall show that in this case every rationally defined divisor class over any field K (containing k) is also rational over K, so that a divisor class $\mathfrak{P}$ is rational over the field $k((G_{\mathfrak{P}}))$. To prove this, let $\mathfrak{o}$ be a divisor of degree 1, rational over k, and let $\mathfrak{O}$ be the class determined by $\mathfrak{o}$. Let $\mathfrak{P}$ be a rationally defined class of degree n over K, and let (p) be a generic point of $\mathfrak{G}_{\mathfrak{P}}$ over K. Then the corresponding positive divisor $\mathfrak{p}$ is rational over $K((p))$, and consequently the class $\mathfrak{P}$ is rational over $K((p))$. There exists an integer r, such that the class $\mathfrak{Q} = \mathfrak{P} - r\mathfrak{O}$ of degree $n - r$ has exactly the dimension 0 and hence contains exactly one positive divisor $\mathfrak{q}$. Since the class $\mathfrak{Q}$ is rational over $K((p))$, the divisor $\mathfrak{q}$ is rational over $K((p))$. Now, let (p') be another generic point of $\mathfrak{G}_{\mathfrak{P}}$ over K, independent with respect to (p) over K; then, by a similar argument, we can conclude that $\mathfrak{q}$ is rational over $K((p'))$. It follows then that the positive divisor $\mathfrak{q}$ is rational over $K((p)) \cap K((p')) = K$, and consequently the class $\mathfrak{Q}$ is rational over K. This implies that the class $\mathfrak{P} = \mathfrak{Q} + r\mathfrak{O}$ is also rational over K.

REFERENCES.

[1] W. L. Chow, "Algebraic systems of positive cycles in an algebraic variety," *American Journal of Mathematics*, vol. 72 (1950), pp. 247-283.

[2] ———, "On compact complex analytic varieties," *American Journal of Mathematics*, vol. 71 (1949), pp. 893-914.

[3] ———, "On the defining field of a divisor in an algebraic variety," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 797-799.

[4] ——— and B. L. van der Waerden, "Zur algebraischen Geometrie. IX. Ueber

zugeordnete Formen und algebraische Systeme von algebraischen Mannigfaltigkeiten," *Mathematische Annalen*, vol. 113 (1937), pp. 692-704.

[5] S. Lefschetz, "On certain numerical invariants of algebraic varieties with application to Abelian varieties," *Transactions of the American Mathematical Society*, vol. 22 (1921), pp. 327-482.

[6] F. K. Schmidt, "Analytische Zahlentheorie in Koerpern der Charakteristik p," *Mathematische Zeitschrift*, vol. 33 (1931), pp. 1-32.

[7] B. L. van der Waerden, "Divisorenklassen in algebraischen Funktionenkoerpern," *Commentarii Mathematici Helvetici*, vol. 20 (1947), pp. 68-109.

[8] A. Weil, *Foundations of Algebraic Geometry*, New York, 1946.

[9] ———, *Varietes abeliennes et courbes algebriques*, Paris, 1948.

[10] O. Zariski, "Foundations of a general theory of birational correspondences," *Transactions of the American Mathematical Society*, vol. 53 (1943), pp. pp. 490-542.

OHIO UNIVERSITY
LIBRARY

LOEWY-GROUPOIDS RELATED TO LINEAR GRAPHS.*

By GERT SABIDUSSI.

1. Introduction. It is not intended to cover much of the theory of L(oewy)-groupoids here. We are content with pointing out a case where L-groupoids appear in the theory of linear graphs.

We begin with a definition: Let G be a finite abstract graph containing no 1-circuit, and let H be a subgraph of G.

Definition. By $M(G, H)$ is denoted the set of all univalent transformations T of H into G which associate i-simplexes of H with i-simplexes of G $(i = 0, 1)$, and preserve the incidence relations of H.

Notations. Throughout this paper any transformation T as defined above will be represented by a transmutation T_0 of the vertices of H together with a transmutation T_1 of the edges of H. The totality of transmutations T_0 for which there is a T_1 such that (T_0, T_1) is an element of $M(G, H)$ will be denoted by $M_0(G, H)$, and the set of those transmutations T_1 for which there exists a T_0 with $(T_0, T_1) \varepsilon M(G, H)$ will be denoted by $M_1(G, H)$.

By $G(H)$ we denote the subgraph of G consisting of those vertices and edges of G which are contained in at least one graph TH, $T \varepsilon M(G, H)$. In symbolic form: $G(H) = M(G, H)H$. Evidently $M(G(H), H) = M(G, H)$.

By $\mathfrak{G}$, $\mathfrak{G}(H)$, $\mathfrak{H}$, etc., we denote the groups of automorphisms of G, $G(H)$, H, etc. We agree that an automorphism of a graph G be represented by a permutation P_0 of the vertices of G together with a permutation P_1 of the edges of G. Here we define $\mathfrak{G}_0$ as the group of those permutations P_0 for which there exists a P_1 with $(P_0, P_1) \varepsilon \mathfrak{G}$, and $\mathfrak{G}_1$ as the group $\{P_1\}$ with $(P_0, P_1) \varepsilon \mathfrak{G}$.

To characterize L-groupoids we shall use the postulates given by A. Loewy [3]:[1]

Given a set $\mathfrak{M}$ of elements which contains a distinguishable subset $\mathfrak{N}$ (the nucleus), and a combination MM' of elements M, M' of $\mathfrak{M}$, $\mathfrak{M}$ is called a *left* L-groupoid, if it satisfies the postulates:

* Received September 28, 1953.

[1] A definition of a similar groupoid was given by R. Baer [1]. Any Baer-groupoid satisfies (1) (2) (3) (4), but not conversely.

(1) $\mathfrak{N}$ is a group under the combination MM'.

(2) MN exists for any $N \in \mathfrak{N}$, $M \in \mathfrak{M} - \mathfrak{N}$, and is a uniquely determined element of $\mathfrak{M} - \mathfrak{N}$.

(3) Let M, M', M'' be elements of $\mathfrak{M}$, then MM' and $(MM')M''$ exist, if $M'M''$ and $M(M'M'')$ exist,[2] and conversely. $M(M'M'') = (MM')M'' = MM'M''$ holds.

(4) There is no $\mathfrak{N}^*$ between $\mathfrak{M}$ and $\mathfrak{N}$ such that the elements of $\mathfrak{N}^*$ and $\mathfrak{M} - \mathfrak{N}^*$ satisfy (1)(2)(3).

If we replace postulates (2)(4) respectively by the postulates

(2′) NM exists for any $N \in \mathfrak{N}$, $M \in \mathfrak{M} - \mathfrak{N}$, and is a uniquely determined element of $\mathfrak{M} - \mathfrak{N}$; and

(4′) There is no $\mathfrak{N}^*$ between $\mathfrak{M}$ and $\mathfrak{N}$ such that the elements of $\mathfrak{N}^*$ and $\mathfrak{M} - \mathfrak{N}^*$ satisfy (1)(2′)(3);

we obtain the set of postulates which define a *right* L-groupoid. Since an L-groupoid has just one nucleus (cf. [3], p. 241, β), $\mathfrak{N}$ may be characterized as the maximal group in $\mathfrak{M}$ such that $\mathfrak{M}$ admits of a decomposition into cosets of $\mathfrak{N}$.

2. The Structure of M, M_0, M_1. The following theorem is fundamental:

THEOREM I. *$M(G, H)$ is a left L-groupoid,[3] $\mathfrak{H}$ being the nucleus of $M(G, H)$.*

As a consequence we have:

THEOREM II. *A necessary and sufficient condition that $M(G, H)$ be a group is that $G(H) = H$.*

Proofs. Theorem I: We prove: A n.a.s.c. that a transformation $T \in M(G, H)$ satisfy $M(G, H)T = M(G, M)$ is that T be an automorphism of H.

This implies that $\mathfrak{H}$ is the maximal group in $M(G, H)$ such that $M(G, H)$ is decomposable into left cosets of $\mathfrak{H}$. Hence $M(G, H)$ is a left L-groupoid, $\mathfrak{H}$ being its nucleus.

---

[2] MM', $(MM')M''$, etc. need not be elements of $\mathfrak{M}$.

[3] More precisely, $M(G, H)$ is a Baer-groupoid.

1. The condition is sufficient. Let $S \varepsilon M(G, H)$, $T \varepsilon \mathfrak{H}$, then $STH = SH$. Hence $ST \varepsilon M(G, H)$, $M(G, H)T = M(G, H)$.

2. The condition is necessary. Let $T = (T_0, T_1)$ be an element of $M(G, H)$ which is not contained in $\mathfrak{H}$. Then T_1 cannot be a permutation. Hence no product ST, $S \varepsilon M(G, H)$, can exist. $M(G, H)T = M(G, H)$ cannot hold.

Theorem II: That the condition is sufficient is evident. The condition is also necessary, since Theorem I implies that: $M(G, H)$ is a group if and only if $M(G, H) = \mathfrak{H}$. But $M(G, H) = \mathfrak{H}$ is equivalent to $G(H) = H$. Theorem II may be given the form: A necessary and sufficient condition that $M(G, H)$ be a group is that G contain no subgraph $H' \neq H$ which is isomorphic with H.

Theorem I implies that $M_0(G, H)$, $M_1(G, H)$ are also L-groupoids. We shall now discuss the relations between the three groupoids M, M_0, M_1. Let G be any graph, then, with the notations as defined in the introduction, and $\mathfrak{G}'$ (a direct product of symmetric groups) being the subgroup of $\mathfrak{G}$ consisting of all elements $(1_0, P_1) \varepsilon \mathfrak{G}$ (1_0 = identity of $\mathfrak{G}_0$), $\mathfrak{G}'$ is a normal subgroup of $\mathfrak{G}$, and $\mathfrak{G}_0 \cong \mathfrak{G}/\mathfrak{G}'$ holds. If G contains no 2-circuit, $\mathfrak{G}'$ reduces to the identity, hence $\mathfrak{G}_0 \cong \mathfrak{G}$. If, moreover, G contains no connected component isomorphic with a single edge incident with two vertices, $\mathfrak{G}_0 \cong \mathfrak{G}_1 \cong \mathfrak{G}$.

One might conjecture that a similar relation holds for $M(G, H)$, $M_0(G, H)$, $M_1(G, H)$, but this is not the case.

Suppose that $G(H)$ contains no 2-circuit, and H contains no connected component isomorphic with the graph H_0 consisting of two vertices joined by an edge. Then given any $T_0 \varepsilon M_0(G, H)$ there is just one $T_1 \varepsilon M_1(G, H)$ such that (T_0, T_1) is an element of $M(G, H)$, and conversely. Thus the following one-one correspondences are set up:

$$M_0(G, H) \longleftrightarrow M(G, H), \qquad M_1(G, H) \longleftrightarrow M(G, H).$$

With respect to these correspondences the behavior of M_0, M_1 is different. We have

THEOREM III. *If $G(H)$ contains no 2-circuit, and H contains no connected component isomorphic with H_0, $M_1(G, H) \cong M(G, H)$ holds;*

whereas the correspondence $M_0 \longleftrightarrow M$ is not necessarily an isomorphism.

Proof of Theorem III. Since M, M_1 are in one-one correspondence, a n.a.s.c. that the two groupoids be isomorphic is that their left nuclei be isomorphic. $\mathfrak{H} \cong \mathfrak{H}_1$, hence $\mathfrak{H}_1$ must be the left nucleus of $M_1(G, H)$. But

this is obvious, for otherwise the nucleus of $M_1(G,H)$ must contain a true transmutation, and hence cannot be a group.

To discuss $M_0(G,H)$ it is useful to introduce subgraphs PH of $G(H)$, $P \varepsilon M_0(G,H)$, where PH means the graph TH, $P \longleftrightarrow T \varepsilon M(G,H)$ under the correspondence $M_0 \longleftrightarrow M$. We may also write $G(H) = M_0(G,H)H$, etc.

Let N be a subgraph of $G(H)$ such that:

1. *N has the same vertices as H;*
2. *$N(H) = N$;*
3. *Any subgraph H' of G which is isomorphic with H is contained in a subgraph N' isomorphic with N;*
4. *$M_0(N,H)$ is a group;*
5. *N is maximal with respect to 1, 2, 3, 4.*

Since H itself satisfies 1, 2, 3, 4 such an N exists. Theorem VI will show that it is unique.

Let $\mathfrak{N}$ be the group of N, then $\mathfrak{N}_0 \cong \mathfrak{N}$, as N is a subgraph of $G(H)$, and $G(H)$ was assumed to contain no 2-circuit.

The above definition of N makes it possible to formulate the following theorems.

THEOREM IV. *If $G(H)$ has the same vertices as H, $M_0(G,H)$ is a right L-groupoid, $\mathfrak{G}_0(H)$ being its nucleus.*

THEOREM V. *If $G(H)$ contains no 2-circuit, $M_0(G,H)$ is a left L-groupoid, $\mathfrak{N}_0$ being its nucleus.*

THEOREM VI. *N is uniquely determined by G, H.*

THEOREM VII. *If $G(H)$ contains no 2-circuit, a necessary and sufficient condition that $M_0(G,H) \cong M(G,H)$ is that $N = H$.*

Proofs. Theorem IV. As $G(H)$ and H have the same vertices, the elements of $\mathfrak{G}_0(H)$ and $M_0(G,H)$ are permutations of the same vertices, and hence may be combined in an arbitrary manner. Obviously $\mathfrak{G}_0(H)$ is a subset of $M_0(G,H)$.

We shall prove: A n.a.s.c. that $T_0 M_0 = M_0$ hold for a permutation of the vertices of $G(H)$ is that $T_0 \varepsilon \mathfrak{G}_0(H)$.

The condition is necessary: Let $S_0 \varepsilon M_0(G,H)$, $T_0 \varepsilon \mathfrak{G}_0(H)$. Then there is an $S_1 \varepsilon M_1$ such that $S = (S_0, S_1) \varepsilon M$, similarly for T_0. Both SH

and TSH are subgraphs of $G(H)$; S, T preserve the incidence relations, hence $T_0S_0 \in M_0(G, H)$.

The condition is also sufficient: Let a' be an edge of $G(H)$. Then, according to the definition of $G(H)$, there is at least one $S \in M(G, H)$ such that SH contains $a' \cdot S = (S_0, S_1) \cdot S_0 = T_0R_0$, $R_0 \in M_0(G, H)$, since $T_0M_0 = M_0$. That means there is an edge a of H such that S_0 maps the end points (P, Q) of a onto the end points (P', Q') of $a' \cdot S_0(P, Q) = T_0R_0(P, Q) = (P', Q')$, $R_0(P, Q) = T_0^{-1}(P', Q')$. R_0P, R_0Q are also vertices of $G(H)$, and since $R_0 \in M_0(G, H)$, there is an $R = (R_0, R_1) \in M(G, H)$, in other words R_0P, R_0Q are joined by an edge of $G(H)$. This means that $T_0^{-1}P'$, $T_0^{-1}Q'$ is an edge of $G(H)$, and since a' was arbitrary, $T_0^{-1} \in \mathfrak{G}_0(H)$. From this, as in the proof of Theorem I, the assertion follows.

Remark. It can be obtained by the same reasoning that $\mathfrak{G}_0$ is a subgroup of $\mathfrak{G}_0(H)$ for any H which is a subgraph of G and has the same vertices as G.

Theorem V: As N has the same vertices as H, the elements of $\mathfrak{H}_0$ and and $M_0(N, H)$ are permutations of the same vertices, and hence can be combined in an arbitrary manner.

We infer from Theorem IV that a n.a.s.c. for $M_0(N, H)$ to be a group is that $M_0(N, H) = \mathfrak{N}_0(H)$. $N(H) = N$ implies $\mathfrak{N}_0(H) = \mathfrak{N}_0$, hence $M_0(N, H) = \mathfrak{N}_0$. Condition 2 of the definition of N implies that any $T \in M_0(G, H)$ can be written in the form $T = SP$ where $P \in M_0(N, H)$, $S \in M_0(G, N)$, hence $M_0(G, H) \subset M_0(G, N)M_0(N, H)$. $M_0(G, H) \supset M_0(G, N)M_0(N, H)$ is obvious, so that we obtain $M_0(G, H) = M_0(G, N)M_0(N, H)$. $M_0(N, H) = \mathfrak{N}_0$ is contained in the left nucleus of $M_0(G, N)$, hence $M_0(G, N) = M_0(G, H)$.

That means $\mathfrak{N}_0$ is contained in the left nucleus of $M_0(G, H)$. We will now prove that $\mathfrak{N}_0$ itself is the left nucleus of $M_0(G, H)$.

Denote, temporarily, by $\mathfrak{N}_0^*$ the left nucleus of $M_0(G, H)$, and suppose that $\mathfrak{N}_0$ is a proper subgroup of $\mathfrak{N}_0^*$.

We denote by N^* the subgraph of $G(H)$ which consists of all subgraphs P_jH, $\mathfrak{N}_0^* = \{P_j\}$, in symbols $N^* = \mathfrak{N}_0^*H$. N^* has the same vertices as H, and it is evident that $N^*(H) = N^*$.

$M_0(G, H)$ may be decomposed into cosets of $\mathfrak{H}_0$, i. e. $M_0(G, H) = \mathfrak{H}_0 + P_2\mathfrak{H}_0 + \cdots + P_p\mathfrak{H}_0$. The cosets $P_i\mathfrak{H}_0$ are in one-one correspondence with those subgraphs of G which are isomorphic with H.

Since $\mathfrak{N}_0^*$ is the left nucleus of $M_0(G, H)$, we have the decomposition $M_0(G, H) = \mathfrak{N}_0^* + R_2\mathfrak{N}_0^* + \cdots + R_r\mathfrak{N}_0^*$, and accordingly $G(H) = \mathfrak{N}_0^*H + R_2\mathfrak{N}_0^*H + \cdots + R_r\mathfrak{N}_0^*H = N^* + R_2N^* + \cdots + R_rN^*$. This means

$G(H)$ is decomposed into a set $\{R_jN^*\}$ of r graphs isomorphic with N^*, no two of which have a common subgraph isomorphic with H. Also condition 2 of the definition of N is satisfied by N^*.

$\mathfrak{N}_0^* = \mathfrak{H}_0 + S_2\mathfrak{H}_0 + \cdots + S_s\mathfrak{H}_0$. Hence N^* contains the graphs H, $S_2H, \cdots, S_sH$. Suppose that N^* contains a graph H' isomorphic with H, but different from the S_iH. In $M_0(G,H)$ H' corresponds to a coset $P\mathfrak{H}_0$ different from the $S_i\mathfrak{H}_0$. $P\mathfrak{H}_0$ must be contained in one of the cosets $R_k\mathfrak{N}_0^*$, hence the graph R_kN^* also contains H'. This is a contradiction, as no two of the graphs R_jN^* have a common subgraph isomorphic with H.

Hence $H, S_2H, \cdots, S_sH$ are the only subgraphs of N^* which are isomorphic with H. But this implies $M_0(N^*,H) = \mathfrak{N}_0^*$. Hence $M_0(N^*,H)$ is a group, i. e. condition 4 of the definition of N is also satisfied by N^*.

$\mathfrak{N}_0 \subset \mathfrak{N}_0^*$, $\mathfrak{N}_0 \neq \mathfrak{N}_0^*$ implies that N is a proper subgraph of N^*. This is a contradiction, since N has been defined as a maximal graph in $G(H)$ satisfying 1234 (This is condition 5). Hence $\mathfrak{N}_0$ is the left nucleus of $M_0(G,H)$.

Theorem VI: $M_0(G,H)$ contains just one left nucleus (cf. [3], p. 241, β), hence $\mathfrak{N}_0 = M_0(N,H)$ is uniquely determined by $M_0(G,H)$. Let N' be a graph which has the same vertices as H, and also satisfies 2345 of the definition of N. Then $M_0(G,N') = M_0(G,H) = M_0(G,N)$. Hence $M_0(N',H) = M_0(N,H)$, and $N' = N'(H) = M_0(N',H)H = M_0(N,H)H = N$, which proves the uniqueness of N.

Theorem VII: Sufficiency is obvious. To prove necessity suppose $H \neq N$. Then $\mathfrak{H}_0$ must be a proper subgroup of $\mathfrak{N}_0$. For $\mathfrak{H}_0 = \mathfrak{N}_0 = M_0(N,H)$ implies $H = \mathfrak{H}_0H = M_0(N,H)H = N$. Hence $\mathfrak{H} \cong \mathfrak{H}_0 \ncong \mathfrak{N}_0$, and $M_0 \ncong M$ follows (cf. [3], p. 245, Satz 1).

3. Conversion theorem.

THEOREM VIII. *Given a finite group $\mathfrak{G}$, and a finite set $\mathfrak{M}_1, \cdots, \mathfrak{M}_r$ of finite left L-groupoids, there always exist cyclically connected graphs $G, H_1, \cdots, H_r$, and connected graphs H_i^k, $i = 1, \cdots, r$, $k = 1, \cdots, s_i$, where s_i is any given positive integer, such that:*

1. *any H_i^k is a subgraph of H_i; any H_i is a subgraph of G;*
2. *H_i^j is a subgraph of H_i^k if and only if $j \leq k$;*
3. *any $M(H_i^k, H_i^j) \cong M_0(H_i^k, H_i^j) \cong M(H_i, H_i^j) \cong M_0(H_i, H_i^j)$, $j \leq k$, is a group.*

4. $\mathfrak{G}$ *is isomorphic with the group of* G;

5. $M(G, H_i^k) \cong M_0(G, H_i^k) \cong M(G, H_i) \cong M_0(G, H_i) \cong \mathfrak{M}_i$,
$i = 1, \cdots, r, k = 1, \cdots, s_i$.

Proof. According to a theorem by R. Frucht [2] one can always find a cyclically connected cubical graph whose group is isomorphic with a given finite group. Denote by $\mathfrak{H}_i$ the nucleus of $\mathfrak{M}_i$, and let H_i', G' be connected cubical graphs whose groups are isomorphic with $\mathfrak{H}_i$, $\mathfrak{G}$ respectively; $i = 1, \cdots, r$. Let H_i'' be the graph which is obtained from H_i' by subdividing every edge of H_i' by i vertices. Thus neither G' nor any H_i'' can contain a subgraph isomorphic with H_j'', $i \neq j$. Obviously H_i'' is connected and has a group isomorphic with $\mathfrak{H}_i$.

We construct a graph H_i^1 by adding to H_i'' five vertices Q_i, R_i^0, R_i^1, R_i^2, S_i joining Q_i with every vertex P_i^j of H_i'', R_i^0 with Q_i, R_i^1 with R_i^0, R_i^2 with R_i^1, and S_i with Q_i, R_i^0, R_i^1, R_i^2. The definition of H_i^k, $k \geq 2$, is, by induction: H_i^k is the graph obtained from H_i^{k-1} by adding a vertex R_i^{k+1} joining it with R_i^k and S_i, $k = 1, \cdots, s_i$. Evidently H_i^j is a subgraph of H_i^k if and only if $j \leq k$.

Now we join S_i with every vertex P_i^j of H_i'', and subdivide the edge $P_i^j S_i$ by j vertices. Since $H_i^{s_i}$ is connected, the graph H_i thus obtained is cyclically connected, and obviously contains all H_i^k, $k = 1, \cdots, s_i$.

Consider $M(H_i^k, H_i^j)$, $j \leq k$. Q_i is the only vertex of H_i^k, H_i^j which is joined with every vertex of H_i'', H_i'' is the only subgraph of H_i^k, H_i^j isomorphic with H_i''; hence any transformation of $M(H_i^k, H_i^j)$ must map H_i'' onto itself, and Q_i onto Q_i. Since S_i is the only vertex of degree $k+3$ of the complement of H_i'' in H_i^k, S_i is mapped onto itself, hence $R_i^{k+1} \to R_i^{k+1}$, and consequently $R_i^j \to R_i^j$, $j \leq k$. In other words $M(H_i^k, H_i^j) \cong$ group of automorphisms of H_i''. If we repeat the above reasoning replacing H_i^k by H_i we obtain $M(H_i, H_i^j) \cong$ group of H_i''. This completes the proof of 123. $M(H_i^k, H_i^j) \cong M_0(H_i^k, H_i^j)$, $M(H_i, H_i^j) \cong M_0(H_i, H_i^j)$ is obvious, since the only subgraph of H_i which has the same vertices as H_i^j, and contains H_i^j is H_i^j itself (cf. the definition of N in section 2).

Take n_i copies $H_{i1}, \cdots, H_{in_i}$ of H_i, where n_i is the order of $\| \mathfrak{M}_i/\mathfrak{H}_i \|$, identify S_{ij} with a vertex of H_{ij+1}'', say P_{ij+1}^1, $j = 1, \cdots, n_i - 1$, finally identify S_{in_i} with P_{i1}^1. The graph H_i^* thus obtained is cyclically connected. To obtain G identify the vertices S_{i1} and S_{i2}, $i = 1, \cdots, r$, and join S_{i1}, S_{i2} with every vertex of G'. Since G' and H_i^* are cyclically connected, so also is G.

Since no H_i is isomorphic with a subgraph of an H_j, $j \neq i$, and no H_i

admits of a "reflection" as automorphism, the group of the complement of G' in G is the identity, if $r \geq 2$, or a cyclic group $\mathfrak{Z}_{n_1}$, if $r = 1$, mapping consecutive subgraphs H_{1j} onto consecutive subgraphs H_{1k} of $H_1{}^*$. This makes it clear that for any $r \geq 1$ the vertices of the complement of G' are fixed vertices of G. Hence the group of $G \cong$ group of $G' \cong \mathfrak{G}$ by construction of G'.

There remains to prove $M(G, H_i) \cong \mathfrak{M}_i$. (That $M(G, H_i) \cong M_0(G, H_i)$ etc., is obvious). From this $M(G, H_i{}^k) \cong \mathfrak{M}_i$ follows directly, since $M(G, H_i{}^k) = M(G, H_i) M(H_i, H_i{}^k)$.

According to a theorem of Loewy (cf. [3], p. 245, Satz 1) two L-groupoids $\mathfrak{M}$, $\mathfrak{M}'$ are isomorphic if and only if their nuclei $\mathfrak{H}$, $\mathfrak{H}'$ are isomorphic, and order $\| \mathfrak{M}/\mathfrak{H} \| =$ order $\| \mathfrak{M}'/\mathfrak{H}' \|$. By the construction of H_i the nucleus of $M(G, H_i)$ is isomorphic with $\mathfrak{H}_i$. As already pointed out in section 2, proof of Theorem V, the elements of the factor-groupoid of $M(G, H_i)$ by its nucleus are in one-one correspondence with those subgraphs of G which are isomorphic with H_i. Hence the order of this factor-groupoid is precisely the number of different graphs isomorphic with H_i. $H_i{}^*$ contains just n_i such graphs, while G', $H_j{}^*$, $j \neq i$, contain none. Hence the order of the factor-groupoid of $M(G, H_i)$ by its nucleus is $n_i =$ order $\| \mathfrak{M}_i/\mathfrak{H}_i \|$, and $M(G, H_i) \cong \mathfrak{M}_i$ follows.

4. Noteworthy special case. In view of Theorem V it would be of interest to have an explicit method of constructing graphs N with $M_0(N, H)$ a group. Unfortunately no such method can be given, except for one trivial case: Let n be the number of vertices of H, then $M_0(C_n, H) = \mathfrak{S}_n$, where C_n is the complete graph on the vertices of H, and $\mathfrak{S}_n$ is the symmetric group on n symbols.

Clearly, for any given special case the graphs N can be computed, which shall be done here for $H = W$ being a *Whitney line* of the given graph G.

Let $\mathfrak{D}_n$ be the dihedral group on n symbols, then clearly the group of automorphisms of an n-circuit (a Whitney line) is isomorphic with $\mathfrak{D}_n$, and will always be referred to as $\mathfrak{D}_n$. The cyclic group $\mathfrak{Z}_n$ on n symbols is a normal subgroup of $\mathfrak{D}_n$, and $\mathfrak{D}_n/\mathfrak{Z}_n \cong \mathfrak{Z}_2$.

THEOREM IX. *$M_0(G, W)$ is either a group, or else a left L-groupoid with nucleus $\mathfrak{D}_n$.*

Let C_n', n even (if n is odd, no such graph C_n' can exist), be the maximal subgraph of C_n which contains no k-circuit, k *odd*. If the labels $1, \cdots, n$ are associated with the vertices of C_n', and $W = (i_1, \cdots, i_n)$ denotes any Whitney line of C_n', C_n' is characterized by the condition that two vertices i_j, i_k

of C_n' are joined by an edge if and only if $j-k$ is odd. The group $\mathfrak{C}_n$ of C_n' consists of those permutations $\left(\begin{smallmatrix}1\cdots n\\ r_1\cdots r_n\end{smallmatrix}\right)$ for which $r_j - r_k \equiv j-k \pmod 2$. $\mathfrak{C}_n \supset \mathfrak{S}_m \times \mathfrak{S}_m \supset \mathfrak{A}_m \times \mathfrak{S}_m \supset \mathfrak{A}_m \times \mathfrak{A}_m \supset \mathfrak{A}_m \supset \mathfrak{E}$, $m = n/2$, $n \geq 10$, is a composition series of $\mathfrak{C}_n$.

THEOREM IX'. *$M_0(G, W)$ is a group in the following cases only:*

$$M_0(G, W) = \mathfrak{D}_n, \text{ if } G(W) = W;$$

$$M_0(G, W) = \mathfrak{S}_n, \text{ if } G = C_n;$$

$$M_0(G, W) = \mathfrak{C}_n, \text{ if } n \text{ is even, and } G(W) = C_n'.$$

Proof of Theorems IX, IX'. Let the notation be so chosen that $W = (1, \cdots, n)$. The totality of permutations P for which $M_0(G, PW) = M_0(G, W)$ is obviously a group; we denote it by $\mathfrak{N}_0 \cdot \mathfrak{N}_0 \subset M_0(G, W)$, $\mathfrak{D}_n \subset \mathfrak{N}_0$.

By R we denote the set of those numbers ρ for which $G(W)$ contains an edge joining the vertices k, $k+\rho$ for some k. Let $P = \left(\begin{smallmatrix}1\cdots n\\ m_1\cdots m_n\end{smallmatrix}\right) \varepsilon \mathfrak{N}_0$, $\rho_i \equiv m_{i+1} - m_i \pmod n$, $1 \leq \rho_i < n$, then $\rho_i \varepsilon R$. Otherwise there must be a k such that the edge joining k, $k+\rho_i$ is not contained in any Whitney line of G, i. e. $Z^{k-m_i} P \notin M_0(G, W)$. ($Z, T$ are the elements which generate $\mathfrak{D}_n \cdot Z^n = 1$, $T^2 = 1$.) This is a contradiction, since Z, P are elements of $\mathfrak{N}_0$.

This makes it evident that $\mathfrak{N}_0 = \mathfrak{D}_n$ and $R = (1, n-1)$ are equivalent statements. Hence we may confine our interest to the case that R contains at least one element ρ, $2 \leq \rho \leq n-2$.

Let $P = \left(\begin{smallmatrix}1\cdots n\\ m_1\cdots m_n\end{smallmatrix}\right) \varepsilon \mathfrak{N}_0, \notin \mathfrak{D}_n$, and ρ_i, $i = 1, \cdots, n$ as defined above. Without loss of generality we may assume that $2 \leq \rho_i \leq [n/2]$ for some i (otherwise take $PT \varepsilon \mathfrak{N}_0$ instead of P). Consider

$$Q_k = Q_k^1 Q_k^2, \cdots, \qquad k = 1, 2, \cdots,$$

$$Q_k^h = (m_i + r + \sigma + h, m_i + r + \sigma + \tau - h + 1), \qquad h = 1, 2, \cdots,$$

where $\rho_i \leq [n/2]$, $r = 1, \cdots, n$, $0 \leq \sigma < \rho_i \leq \sigma + \tau \leq n-1$, $\tau = \rho_i$ if $k = 1$; $\tau \equiv \rho_i \pmod 2$, $1 \leq \tau \leq k(\rho_i - 1) + 1$, if $k \geq 2$. The difference of the numbers s_j, s_{j+1} into which $j, j+1$ ($j = 1, \cdots, n$) are carried by Q_k is congruent to 1, $n-1$, or τ. Hence $Q_1 \varepsilon M_0(G, W)$; moreover, as we shall see below, $Q_k \varepsilon M_0(G, W)$ implies $Q_{k+1} \varepsilon M_0(G, W)$. Suppose $Q_k \varepsilon M_0(G, W)$. $Q_k Z^r P \varepsilon M_0(G, W)$ carries over

$$i \to m_i + r, \qquad i+1 \to m_i + r - \rho_i + 2\sigma + \tau + 1, \qquad (r = 1, \cdots, n).$$

Hence $\tau^* = 2\sigma - \rho_i + \tau + 1 \varepsilon R$. If $\rho_i \geqq (n+k-1)(k+1)^{-1}$, τ^* ranges

over all odd or even numbers between 1 and $n-1$, according as ρ_i is odd or even. If $\rho_i < (n+k-1)(k+1)^{-1}$ the maximum value attained by τ^* under the above conditions for σ, τ is $(k+1)(\rho_i - 1) + 1$ which is precisely the upper bound for τ in the conditions for Q_{k+1}. Thus $Q_{k+1} \varepsilon M_0(G, W)$ is ascertained, and the whole reasoning may be repeated with Q_k replaced by Q_{k+1}.

After a certain number of steps we obtain that any $\tau \equiv \rho_i \pmod 2$, $2 \leq \tau \leq n-2$, is an element of R. If n is odd, any $\tau' \not\equiv \rho_i \pmod 2$, $2 \leq \tau' \leq n-2$, is also contained in R, since $\tau \varepsilon R$ implies $n - \tau \varepsilon R$. This means $R = (1, \cdots, n-1)$, $G(W) = G = C_n$.

If n is even, any $P \varepsilon \mathfrak{N}_0$ must have at least one odd ρ_i, which implies $1, 3, 5, \cdots, n-3, n-1 \varepsilon R$, i. e. C_n' is a subgraph of $G(W)$. That means $M_0(G, W)$ is an L-groupoid whose nucleus is either $\mathfrak{D}_n$ or $\mathfrak{C}_n$. Assume the latter, and $G(W) \neq C_n'$. Then $G(W)$ must contain a Whitney line containing two edges $(r-a, r+a)$, $(s-b, s+b)$ which are not contained in C_n'.

Let $W' = (i_1', \cdots, i_n')$ be a Whitney line of C_n', $n \geq 4$, $W = (i_1, \cdots, i_m)$ a Whitney line of C_m, $m = n/2$, then we denote by C_n'' the graph which is obtained by identifying the vertices i_j and i_{2j}', $j = 1, \cdots, m$, of C_m and C_n'. Clearly C_n'' is independent of W, W', and $C_n''(W) = C_n'$ holds (this can easily be proved by induction). $C_n''(W) = C_n'$ implies $H(W) = C_n'$ for any subgraph H of C_n'' which contains C_n'. In other words, let $(r-a, r+a)$, $(s-b, s+b)$ be the edges introduced above, then $r \pm s \not\equiv a \pm b \pmod 2$. Hence $G(W)$ contains a Whitney line in which $(r-a-1, r+a)$, $(r+a, r-a)$, $(r-a, r-a+1)$ are consecutive edges; i. e. there is a $P \varepsilon M_0(G, W)$ carrying over $r-a-1 \to r-a-1$, $r-a+2 \to r-a+1$.

$$Q = (r-a, r-a+2)(r-a+1, r-a+3) \varepsilon \mathfrak{C}_n,$$

hence $PQ \varepsilon M_0(G, W)$. PQ carries over

$$r-a-1 \to r-a-1, \qquad r-a \to r-a+1.$$

Consequently the vertices $r-a-1$, $r-a+1$ are joined in $G(W)$. So are $s-b-1$, $s-b+1$ which follows readily by interchanging the roles of r, a and s, b.

Let the vertices $r-1$, $r+1$ and $s-1$, $s+1$, $r-s$ odd, be joined in $G(W)$. Then $P = (r, r+1, \cdots, s-3, s-2)(s-1, s, s+1) \varepsilon M_0(G, W)$. $Q = (s, s+2)(s+1, s+3) \varepsilon \mathfrak{C}_n$, hence $PQ \varepsilon M_0(G, W)$. PQ carries over: $s-1 \to s$, $s \to s+2$. Hence the vertices s and $s+2$ are joined in $G(W)$.

So are r, $r+2$. The reasoning may be repeated with $r-1$, $r+1$ replaced by r, $r+2$, etc., and we obtain $2 \in R$.

Since $2 \in R$,

$$P = (k, k+1, \cdots, k+m-2)(k+m-1, k+m, k+m+1) \in M_0(G, W).$$

By P the vertices of $G(W)$ are divided into two classes

$$S_k = (k, k+1, \cdots, k+m-2, k+m, k+m+1),$$

where $m = n/2$, $S_k' =$ complement of S_k in $(1, 2, \cdots, n)$ such that any vertex $r \in S_k$ can be reached from any vertex $s \in S_k'$, $r-s$ even, by passing through an odd number of edges of PW. This means there is a suitable $Q \in \mathfrak{C}_n$ such that r and s are consecutive vertices of the Whitney line PQW. Hence $r, s, r-s$ even, are joined in $G(W)$. If we take $k = 1, 2, \cdots, n$, we obtain that any even number is contained in R, which together with $1, 3, \cdots, n-3, n-1 \in R$ implies $G(W) = G = C_n$, a contradiction, since G was assumed to be a proper subgraph of C_n.

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] R. Baer, "Zur Einordnung der Theorie der Mischgruppen in die Gruppentheorie," *Sitzungsberichte der Heidelberger Akademie der Wissenschaften*, Jahrgang 1928, 4. Abhandlung.

[2] R. Frucht, "Graphs of degree 3 with given abstract group," *Canadian Journal of Mathematics*, vol. 1 (1949), pp. 365-378.

[3] A. Loewy, "Ueber abstrakt definierte Transmutationssysteme oder Mischgruppen," *Journal fuer die reine und angewandte Mathematik*, vol. 157 (1927), pp. 239-254.

THE INSOLUBILITY OF CLASSES OF DIOPHANTINE EQUATIONS.*

By N. C. ANKENY and P. ERDÖS.

Introduction. Consider the non-trivial rational integer solutions in the variables $X_1, X_2, \cdots, X_n$ of the equation

$$a_1X_1^m + a_2X_2^m + \cdots + a_nX_n^m = 0, \tag{1}$$

where $m, a_1, a_2, \cdots, a_n$ are non-zero rational integers, and $m > 0$. By a non-trivial solution we mean one in which not all $X_j = 0$, $j = 1, 2, \cdots, n$.

Let U be a large positive real number tending to infinity, and let $D(U, a_1, a_2, \cdots, a_n) = D(U)$ be the number of $m \leqq U$ for which (1) has a non-trivial rational solution. Putting a mild but necessary restriction on the coefficients, something may be said about the order of magnitude of $D(U)$.

THEOREM I. *If, for every selection of* $e_j = 0$ *or* ± 1, $(j = 1, 2, \cdots, n)$ *except* $(e_1, \cdots, e_n) = (0, 0, \cdots, 0)$, *we have* $a_1e_1 + \cdots + a_ne_n \neq 0$, *then* $D(U) = o(U)$ *as* $U \to +\infty$.

Theorem I could be interpreted as stating that equation (1) is "almost always" unsolvable; or the density of m, for which (1) has a non-trivial solution, is zero.

One very important case that the hypothesis of Theorem I excludes is when $a_1 = a_2 = a_3 = 1$. However, our methods still yield a result of some interest in this case.

THEOREM II. *The density of integers* m, *for which the equation* $X_1^m + X_2^m + X_3^m = 0$ *has a rational solution and for which* $(X_1X_2X_3, m) = 1$, *is zero.*

The restriction $(X_1X_2X_3, m) = 1$ is sometimes referred to as the first case in Fermat's equation.

The result $M(U) = o(U)$ can be strengthened to $M(U) = O(U(\log U)^{-c})$ for some positive constant c. The proof of this stronger inequality requires a good deal more effort and will not be presented in this paper.

The result of Theorem I can be generalized from the rational number

* Received September 14, 1953; revised December 7, 1953.

field to any algebraic number field F. The restriction on a_j, which are now any non-zero algebraic integers contained in F, is that $a_1\epsilon_1 + \cdots + a_n\epsilon_n \neq 0$ where $\epsilon_j = 0$ or any root of unity contained in F. The proof of this generalization will not be given in complete detail, but will be briefly outlined at the end of this paper.

In Section 1 we shall present some introductory Lemmas and in Section 2, the proof of Theorems I and II will be presented.

1. *Notations.* U denotes a large positive variable. $c_1, c_2, \cdots$ denote absolute constants. p, q are rational primes. ζ_g is a primitive g-th root of unity.

LEMMA 1. *Let $a_1, \cdots, a_n$ satisfy (2), $g > 2$, and $4 \nmid g$. If $(e_1, \cdots, e_n)$ is any one of the $3^n - 1$ n-tuples referred to in the statement of Theorem* I *and if $h_1, \cdots, h_n$ are any non-negative integers then*

$$\sum_{k=1}^{n} a_k e_k \zeta_g^{h_k} \neq 0. \tag{3}$$

Proof. Suppose first that $g = p$ or $2p$ where p is an odd prime. Since $\zeta_g^p = \pm 1$, the assumption that (3) is false leads to a relation

$$\sum_{j=0}^{p-1} b_j \zeta_g^j = 0, \text{ where } b_j = \sum_{k \in S_j} a_k e_k, \qquad (j = 0, \cdots, p-1),$$

and S_j is a (possibly void) subset of the set of numbers $\{1, \cdots, n\}$. The sets $S_0, \cdots, S_{p-1}$ are non-overlapping and their union is the set $\{1, \cdots, n\}$. Thus, because of (2), there is an i such that $b_i \neq 0$ and, for every $i' \neq i$, $b_{i'} \neq b_i$. On the other hand, ζ_g is a root of either $x^{p-1} + x^{p-2} + \cdots + x + 1$ or $x^{p-1} - x^{p-2} + - \cdots + (-1)^{p-1}$, both of which are irreducible polynomials over the rational field R. It follows that $b_0 = \pm b_1 = \cdots = \pm b_{p-1}$, a contradiction.

To complete the proof of the lemma, let $g = p_1^{d_1} p_2^{d_2} \cdots p_s^{d_s}$ or $2p_1^{d_1} p_2^{d_2} \cdots p_s^{d_s}$ where the p's are distinct odd primes and the d's are positive integers. Assume by induction on $d_1 + \cdots + d_s$ that the lemma holds for $g' = g/p_1$ (> 2). Since $\zeta_g^{p_1} = \zeta_{g'}$ the assumption that (3) is false leads to a relation

$$\sum_{j=0}^{p_1-1} \beta_j \zeta_g^j = 0, \text{ where } \beta_j = \sum_{k \in S_j} a_k e_k \zeta_{g'}, \qquad (j = 0, \cdots, p_1 - 1),$$

the f's being non-negative integers and the sets $S_0, \cdots, S_{p_1-1}$ having a meaning similar to that in the first part of the proof.

By the inductive hypothesis there is an i such that $\beta_i \neq 0$ and, for every

$i' \neq i$, $\beta_{i'} \neq \beta_i$. On the other hand, the irreducible equation satisfied by ζ_g in the field $R(\zeta_{g'})$ is either $x^{p_1-1} + \cdots + x + 1 = 0$, $(d_1 = 1)$, or $x^{p_1} - \zeta_{g'} = 0$, $(d_1 > 1)$. Thus $\beta_0 = \beta_1 = \cdots = \beta_{p_1-1}$, a contradiction.

LEMMA 2. *If* $3 \nmid g$ *then, for any non-negative integers* h_1, h_2, h_3,

$$\zeta_g^{h_1} + \zeta_g^{h_2} + \zeta_g^{h_3} \neq 0. \tag{4}$$

Proof. Assume there exist h_1, h_2, h_3 such that $\zeta_g^{h_1} + \zeta_g^{h_2} + \zeta_g^{h_3} = 0$. Divide through by $\zeta_g^{h_3}$, yielding

$$\zeta_g^{k_1} + \zeta_g^{k_2} + 1 = 0, \tag{5}$$

for 2 integers k_1, k_2. Taking the imaginary parts of both sides of (5) yield that $\sin(2\pi k_1/g) + \sin(2\pi k_2/g) = 0$. This implies $k_1 \equiv -k$, or $k_1 + g/2 \pmod g$ where only the former is possible if $2 \nmid g$.

Now taking the real part of (5) yields $\cos(2\pi k_1/g) + \cos(2\pi k_2/g) = -1$ or, on substituting $k_2 \equiv -k_1$ or $k_1 + g/2 \pmod g$, yields that

$$2\cos(2\pi k_1/g) = -1, \text{ or } \cos(2\pi k_1/g) + \cos(2\pi(k_1 + 2g/g)) = -1.$$

This last equation is clearly impossible. The former equation implies that $3 \mid g$, which is contrary to our hypothesis.

THEOREM III. *If* $a_1, a_2, \cdots, a_n$ *satisfy condition* (2), *then for a given* m *there exists no non-trivial rational solutions of* (1) *provided we can find a rational prime* p *such that*

$$m \text{ divides } p - 1, \qquad mr = p - 1, \tag{6}$$

$$4 \nmid r \tag{7}$$

$$\phi(r) < \alpha^{-1} \log p, \tag{8}$$

where $\alpha = \log(|a_1| + |a_2| + \cdots + |a_n|)$, *and* $\phi(r)$ *is a Euler* ϕ *function.*

Proof. (cf. [4], H. S. Vandiver). Assume there exists a p which satisfies (3), (4) and (5), and that (1) has a rational solution such that $X_1 X_2 \cdots X_n \not\equiv 0 \pmod p$. Without loss of generality, assume $(X_1, X_2, \cdots, X_n) = 1$. Then consider (1) in the field $R(\zeta_r)$.

As $p \equiv 1 \pmod r$, the ideal factorization of p is $(p) = P_1 P_2 \cdots P_s$ in $R(\zeta_r)$, where $s = \phi(r)$, and $N_{R(\zeta_r),R}(P_1) = p$. Hence, the group of m-th power residues of the multiplicative cyclic group of residues $(\bmod P_1)$ has $(p-1)/m = r$ elements. One sees that the elements ζ_r^j, $j = 0, 1, \cdots, r-1$ are incongruent $(\bmod P_1)$. So ζ_r^j form a subgroup of r elements in a multi-

plicative subgroup of residues (mod P_1). Hence, these two subgroups must coincide.

As $a_1X_1^m + \cdots + a_nX_n^m = 0$, à fortiori, $a_1X_1^m + \cdots + a_nX_n^m \equiv 0 \pmod{P_1}$ or, by the coinciding of the two subgroups, $a_1\zeta_r^{t_1} + \cdots + a_n\zeta_r^{t_n} \equiv 0 \pmod{P_1}$ for some n-tuple of integers $(t_1, \cdots, t_n)$. Hence, $p = N_{R(\zeta_r),R}(P_1)$ divides $N_{R(\zeta_r),R}(a_1\zeta_r^{t_1} + \cdots + a_n\zeta_r^{t_n})$. But,

$$| N_{R(\zeta_r),R}(a_1\zeta_r^{t_1} + \cdots + a_n\zeta_r^{t_n}) \leqq (| a_1 | + \cdots + | a_n |)^{\phi(r)}.$$

Thus $p \leqq (| a_1 | + \cdots + | a_n |)^{\phi(r)}$, which is a contradiction to our hypothesis unless $a_1\zeta_r^{t_1} + \cdots + a_n\zeta_r^{t_n} = 0$. This case, however, is impossible by Lemma 1.

Hence, we have shown that $X_1X_2 \cdots X_n \equiv 0 \pmod{p}$. Hence, p divides one of the variables, say X_n. However, proceeding in the same way with the truncated equation $a_1X_1^m + a_2X_2^m + \cdots + a_{n-1}X_{n-1}^m$ we will see that p will divide each X_i, $i = 1, 2, \cdots, n$. This is a contradiction to $(X_1, X_2, \cdots, X_n) = 1$. This proves Theorem 1.

COROLLARY. *If $n = 3$, $a_1 = a_2 = a_3 = 1$, m square free, and a prime p exists that satisfies* (6), (7), (8) *in Theorem* III, $3 \nmid r$, *then* (1) *has no non-trivial solution relatively prime to m.*

Proof. Using the proof of Theorem III and Lemma 2, we immediately infer that there exists no solution of $X_1^m + X_2^m + X_3^m \equiv 0 \pmod{p}$ and $X_1X_2X_3 \not\equiv 0 \pmod{p}$. Hence, if there exists a rational solution $X_1^m + X_2^m + X_3^m = 0$, then $p \mid X_1X_2X_3$.

If q denotes any prime factor of m, and $(X_1X_2X_3, m) = 1$ we have, by using Furtwangler's criterion on Fermat's Equation (cf. Landau [2]), that for any $p \mid X_1X_2X_3$, $p^{q-1} \equiv 1 \pmod{q^2}$. As $p \equiv 1 \pmod{m}$, $p \equiv 1 \pmod{q}$. Therefore, $p \equiv 1 \pmod{q^2}$. As m is square free, $p \equiv 1 \pmod{m^2}$; therefore $p - 1 \geqq m^2$.

By hypothesis, $\phi(r) < \log p/\log 3$. Thus $r < (\log p/\log 3)^2$. Now $m^2 \leqq p - 1 = mr < (\log p/\log 3)^2 m$. Hence, $m < (\log p/\log 3)^2$ or $p - 1 < (\log p/\log 3)^4$.

This last inequality is clearly contradictory and this completes the proof of the corollary.

2. To prove Theorems I and II, we shall derive a set of integers m which satisfy Theorem III and such that almost all integers are divisible by at least one element of our set.

Denote by $\lambda(n)$ the least prime diivsor of n.

LEMMA 3. *If γ denotes Euler's constant,*

$$\sum_{n \in J_1(U)} 1 = e^{-\gamma} U (\log\log\log U)^{-1} + O(\log U),$$

where $J_1(U)$ denotes the rational integers lying between U and $2U$ which have all their prime factors $> \log\log U$.

LEMMA 4. *If $d < U^{\frac{1}{2}}$, then*

$$\sum_{n \in J_2(U)} 1 = e^{-\gamma} \phi(d, \log\log U) U (\log\log\log U)^{-1} + O(\log U)$$

where $\phi(d, V) = d \prod_{\substack{p \mid d \\ p < V}} (1 - 1/p)$, and $J_2(U)$ is the set of integers n between U and $2U$, and $n \equiv 1 \pmod{d}$.

LEMMA 5. *For any constant c_1, and U sufficiently large,*

$$\sum_{\log U < r < 2\log U} \sum_{J_3(U,r)} 1 > c_2 U (\log\log\log U)^{-1},$$

where $J_3(U, r)$ denotes the set of primes $p < c_1 U \log U$, $p \equiv 1 \pmod{r}$, and $\lambda((p-1)/r) > \log\log U$. The constant c_2 depends only upon the choice of c_1.

Lemmas 3, 4, and 5 are quite elementary in nature. The proofs of them are very similar. We shall give here only a proof of Lemma 3.

Proof of Lemma 3. Let d be any square free number $< \log\log U$, and let $f(d, U)$ denote the number of integers which lie between U and $2U$ and which are divisible by d. Then $f(d, U) = U/d + O(1)$. If $\mu(d)$ denotes the Moebius function

$$\sum_{n \in J_1(U)} 1 = \sum_{U < n < 2U} \sum_{d \mid (n, h)} \mu(d)$$

where $h = \prod_{p \leqq \log\log U} p$, as this last inner sum is 1 if n has no prime factors $\leqq \log\log U$, and zero otherwise. Hence,

$$\sum_{n \in J_1(U)} 1 = \sum_{d \mid h} \mu(d) \sum_{\substack{U < n < 2U \\ d \mid n}} 1 = \sum_{d \mid h} \mu(d) f(d, U)$$

$$= \sum_{d \mid h} (\mu(d) U/d + O(1)) = U \prod_{p \mid h} \mu(d)/d + O\Big(\sum_{d \mid h} 1\Big)$$

$$= U \prod_{p \mid h} (1 - 1/p) + O(h) = e^{-\gamma} U (\log\log\log U)^{-1} + O(\log U)$$

by using Merton's Theorem on prime numbers. This proves Lemma 3. The proof of Lemma 4 is almost identical to the above. Lemma 5 is again the same as above using the Siegel-Walficz theorem on primes in arithmetic progressions.

Lemma 4 implies that to the primes $< c_3 U \log U$ there corresponds at least $c_2 U(\log\log\log U)^{-1}$ numbers $m < U$, $m = (p-1)/r$, $\lambda(m) > \log\log U$. However, this correspondence is not necessarily unique, and possibly many primes could correspond to the same m. With this in mind we prove

LEMMA 6. *Let $H(U, g)$ denote the number of integers m, $U < m < 2U$, $\lambda(m) > \log\log U$, and such that there are exactly g distinct primes $p_1, p_2, \cdots, p_g$ where $p_j \equiv 1 \pmod m$, and $p_j < c_1 U \log U$ for $j = 1, 2, \cdots, g$. Then $H(U, g) = O(g^{-2} U \log\log\log U)$.*

Proof. The function g^{-2} in the above Lemma could easily be replaced by a function of g which tends to zero far more rapidly as g increases, but this improvement is not needed for this present paper.

To prove Lemma 6 we shall derive an upper bound on the number of times that $(r_1 m + 1)$, $(mr_2 + 1)$, $(mr_3 + 1)$ can simultaneously be primes where $\lambda(m) > \log\log U$, $U \leq m \leq 2U$, $1 \leq r_1, r_2, r_3 \leq \log U$.

Now for the moment regard r_1, r_2, r_3 as fixed and m varying as we described. Then the problem is to derive an upper bound on the number of elements

$$(mr_1 + 1)(mr_2 + 1)(mr_3 + 1) \tag{9}$$

which have no prime factors $\leq U^{\frac{1}{3}}$. This corresponds to the slight generalization to the twin prime problem where there the number we sieve is $(m)(m+2)$. In our problem we have a polynomial in m composed of 3 linear factors. Utilizing the general method developed by Selberg (cf. [3], especially pp. 291-292), we can easily prove that there are less than

$$c_4 U(\log\log\log U)^{-1}(\log U)^{-3}\psi(r_1)\psi(r_2)\psi(r_3)$$

element of (9) with r_1, r_2, r_3 fixed which have no prime factors $< U^{\frac{1}{3}}$, where $\psi_3(r) = \prod_{p \mid r}(1 - 1/p)$.

Hence, summing over r_1, r_2, r_3, we have that the number of elements of (9) is less than $c_4 U(\log\log\log U)^{-1}(\log U)^{-3}(\sum_{r=1}^{\log U}\psi(r))^3$. Now

$$\psi(r) = \prod_{p|r} (1 - 1/p)^{-1}$$
$$= \prod_{p|r} (1 + 1/p)(1 - 1/p^2)^{-1} < \prod_{p|r} (1 + 1/p) \prod_{p} (1 - 1/p^2)^{-1}$$
$$\leqq \pi^2/6 \prod_{p|r} (1 + 1/p) = c_5 \sum_{d|r} 1/d.$$

Hence

$$\sum_{r=1}^{\log U} \psi(r) < c_5 \sum_{r=1}^{\log U} \sum_{d|r} 1/d = c_5 \sum_{d=1}^{\log U} 1/d \sum_{\substack{d|r \\ r<\log U}} 1 \leqq c_5 \sum_{d=1}^{\log U} \log U/d^2 < c_6 \log U.$$

Hence, the number of elements of (9) which have all their prime factors $< U^{\frac{1}{2}}$ is $< c_7 U (\log\log\log U)^{-1}$. However, if $c_m{}^n$ is the binomial coefficient, this number is equal to $\sum_{g=3}^{\infty} c_3{}^g H(U, g)$ where $g > 2$. Therefore

$$\sum_{g=3}^{\infty} c_3{}^g H(U, g) < c_6 U (\log\log U)^{-1}. \tag{10}$$

Lemma 6 follows immediately from (10).

Lemmas (4) and (5) showed that to the primes corresponded various numbers m. Lemma 6 shows conversely that to each m there cannot correspond too many primes.

Therefore, if we define the set $M(U)$ to be all integers $m < U$, $\lambda(m) > \log\log U$, and m satisfies the conditions of Theorem III, then

$$\sum_{m \in M(U)} 1 > c_7 U (\log\log\log U)^{-1}. \tag{11}$$

3. In this section we shall establish that almost all rational integers are divisible by an integer of the set M where $M = \bigcup_{c_8 < U < \infty} M(U)$.

Let $D(M(U)) = D(U)$ denote the density of integers not divisible by any integer of our set $M(U)$. Let U_1 be some large real number.

LEMMA 7. *There exists a constant* $0 < c_8 < 1$ *such that* $D(U_1) < c_8$.

Proof. As $D(U_1)$ denotes the density of integers not divisible by any $m \in M(U_1)$, $1 - D(U_1)$ denotes the density of integers which are divisible by some $m \in M(U_1)$. Hence

$$1 - D(U_1) \geqq \sum_{m \in M(U_1)} \delta(m), \tag{12}$$

where $\delta(m)$ denotes the density of integers divisible by m but by no other integer of our set $M(U_1)$ except a divisor of m which might be contained in $M(U_1)$. Now, the density of integers t which have no prime factors between

$\log\log U_1$ and U_1, is well known to be $> \frac{1}{2}(\log\log\log U_1)(\log U_1)^{-1}$. Now the set mt is divisible by m and no other element of our set $M(U_1)$ except possibly a divisor of m, as all numbers of $M(U_1)$ have all of their prime divisors lying between $\log\log U_1$ and U_1, and hence to divide mt implies, by our definition of the integers t, that it would divide m. Therefore,

$$\delta(m) \geqq \tfrac{1}{2}(\log\log\log U_1)(\log U_1)^{-1}m^{-1}. \tag{13}$$

By (12), (13) and (11),

$$1 - D(U_1) \geqq \tfrac{1}{2}(\log\log\log U_1)(\log U_1)^{-1} \sum_{m \in M(U_1)} 1/m > c_9 > 0.$$

Letting $c_8 = 1 - c_9$, we have proved Lemma 7.

If U_2 is another large constant, then by Lemma 7, $D(U_2) < c_8$ also. If $U_2 > \exp\{\exp\{2U_1\}\}$, then the elements of $M(U_1)$ are relatively prime to the elements of $M(U_2)$. As all prime factors of $M(U_2)$ are greater than $\log\log U_2 > 2U_1$, and all prime factors of the elements of $M(U_1)$ are less than U_1.

As $D(M(U))$ denotes the density of integers not divisible by any element in $M(U)$,

$$D(M(U_1) \cup D(M(U_2))) = D(M(U_1)) \cdot D(M(U_2)) < c_8^2.$$

Similarly, defining $U_3 = \exp\{\exp\{2U_2\}\}$, $U_4 = \exp\{\exp\{2U_3\}\}, \cdots$, gives that

$$D(\bigcup_{j=1}^{n} M(U_j)) = \prod_{j=1}^{n} D(M(U_j)) < c_8^n \to 0$$

as $n \to \infty$. Hence,

$$D(M) \leqq \lim_{n\to\infty} D(\bigcup_{j=1}^{n} M(U_j)) = 0, \tag{14}$$

where $D(M)$ denotes the density of integers not divisible by any element of M.

Conversely (14) may be interpreted as saying that almost all integers are divisible by some element of our set M. If an integer n is divisible by an m, $m \in M$, we see that the equation (1) has no non-trivial solution for m, and hence, no non-trivial solution for n. This completes the proof of Theorem I.

To prove Theorem II we would need to add to our conditions on M that the $(p-1)/m$ be relatively prime to 3, and that m be square free. These additional assumptions could easily be incorporated in Section II, and present no real difficulties.

To establish the generalization of Theorem I to an algebraic number

field F we merely sum in Lemma 5 over the rational primes which are norms of prime ideals in F. Theorem III can be bodily carried over by changing the definition of α to $(F:R)\log(|a_1|+\cdots+|a_n|)$ where $(F:R)$ denotes the degree of F over the rational number. The remainder of the proof is almost identical.

THE JOHNS HOPKINS UNIVERSITY
AND
AMERICAN UNIVERSITY.

REFERENCES.

[1] N. C. Ankeny, "The insolubility of sets of diophantine equations in the rational numbers," *Proceedings of the National Academy of Sciences*, vol. 38 (1952), pp. 880-884.

[2] E. Landau, *Vorlesungen über Zahlentheorie*, vol. III, 1950, p. 315.

[3] A. Selberg, "Sieve-method in prime number theory," *Proceedings of the international Congress of Mathematicians*, vol. I (1950), pp. 286-292.

[4] H. S. Vandiver, "On classes of diophantine equations of higher degrees which have no solutions," *Proceedings of the National Academy of Sciences*, vol. 32 (1946), pp. 101-106.

ON CURVES DEFINED BY BINARY NON-CONSERVATIVE DIFFERENTIAL SYSTEMS.*

By PHILIP HARTMAN and AUREL WINTNER.

If a prime denotes d/dt and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{1}$$

is a constant matrix, then the determination of the solution paths $(x(t), y(t))$ of

$$x' = ax + by, \qquad y' = cx + dy \tag{2}$$

in the real (x, y)-plane, hence the description of their asymptotic behavior as $t \to \infty$, is elementary. From the second of these points of view, the so-called geometrical theory of real, binary differential systems, as initiated (in the case of power series f, g) by Poincaré and extended (to non-analytic cases) by Bendixson and Perron, deals with conservative but non-linear "small perturbations" of (2), namely, with a system of the form

$$x' = ax + by + f(x, y), \qquad y' = cx + dy + g(x, y) \tag{3}$$

where both functions f, g are real-valued and continuous for small

$$r = (x^2 + y^2)^{\frac{1}{2}} \geqq 0 \tag{4}$$

and satisfy

$$f(x, y)/r \to 0, \qquad g(x, y)/r \to 0 \text{ as } r \to 0, \tag{5}$$

while

$$ad - bc \neq 0. \tag{6}$$

But there is another standard "small perturbation" of (2), one which is non-conservative but linear and homogeneous, that is, of the form

$$x' = \alpha(t)x + \beta(t)y, \qquad y' = \gamma(t)x + \delta(t)y, \tag{7}$$

where α, β, γ, δ are real, continuous functions of t satisfying

$$|\alpha(t) - a| + |\beta(t) - b| + |\gamma(t) - c| + |\delta(t) - d| \to 0 \text{ as } t \to \infty. \tag{8}$$

* Received November 16, 1953.

In [2], known results on the asymptotic behavior of the solution paths $(x(t), y(t))$ reaching to the origin, that is, satisfying

$$r(t) \to 0 \quad \text{as} \quad t \to \infty, \tag{9}$$

were collected and refined for the problem (3)-(5). The present paper will transfer those considerations from (3)-(5) to (7)-(8) and to certain generalizations of both (7)-(8) *and* (3)-(5); cf. Theorems (*) *and* (†) below.

A question considered in [2] for (3)-(5) is whether or not a solution path $(x(t), y(t))$ reaching to the origin (as $t \to \infty$) has a tangent there, that is, whether or not a continuous determination of $\theta(t)$, where

$$\theta = \arctan y/x, \tag{10}$$

is such that

$$\lim_{t \to \infty} \theta(t) = \theta_0 \text{ exists} \tag{11}$$

(as a finite limit). For the case in which such a tangent exists, there was considered in [2] the question whether or not the half-line $\theta = \theta_0$ is the limit of tangents, that is, whether or not $x'^2 + y'^2 \neq 0$ holds for large t and a continuous determination of $\psi(t)$, where

$$\psi(t) = \arctan y'/x', \tag{12}$$

is such that

$$\lim_{t \to \infty} \psi(t) \text{ exists and is } \theta_0 (\mathrm{mod}\, \pi). \tag{13}$$

In the case of the linear system (7)-(8), these questions concerning (10), (11) and (12), (13) can be raised for all solution paths (and not only for those reaching to the origin). In view of (4), the differential equation satisfied by (10) is

$$\theta' = \gamma(t)\cos^2\theta + (\delta(t) - \alpha(t))\cos\theta \sin\theta - \beta(t)\sin^2\theta. \tag{14}$$

Thus it cannot depend on (9) whether (11) does or does not hold; in fact, if a sufficiently large constant is subtracted from both α and δ, then (9) holds for all solutions and (14) remains unchanged.

The answer to the question concerning (12), (13) for the linear system (7)-(8) is trivially in the affirmative when (6) holds.

Consider the following classification of the possible matrices (1): The matrix (1) has

(i) real, distinct characteristic numbers;

(ii) non-real characteristic numbers;

(iii) equal characteristic numbers, multiple elementary divisors;
(iv) equal characteristic numbers, simple elementary divisors.

There will be no statements concerning the case (iv).

(*) *Let* $\alpha(t)$, $\beta(t)$, $\gamma(t)$, $\delta(t)$ *be continuous functions on* $0 \leqq t < \infty$ *satisfying* (8). *In case* (i), *every solution* $\theta = \theta(t)$ *of* (14) *satisfies* (11); *in case* (ii), *every solution* $\theta = \theta(t)$ *satisfies*

$$|\theta(t)| \to \infty \quad \text{as} \quad t \to \infty; \tag{15}$$

in case (iii), *either all solutions satisfy* (11) *or all satisfy* (15).

Furthermore, in the cases (i) and (iii), the only possible limits (11) are the zeros of

$$c\cos^2\theta + \tfrac{1}{2}(d-a)\sin^2\theta - b\sin^2\theta = 0. \tag{16}$$

In case (i), each of the two distinct zeros (mod π) is the limit (11) for some solution; in case (iii), (16) has only one zero (mod π).

The assertion of (*) concerning (ii) is trivial. That concerning (i) is contained in sharper results of Perron [3], pp. 148-160. It will be clear from the proof of (*) that both alternatives can be realized in the case (iii).

Proof of (*). Although the assertions of (*) concern merely the solutions $\theta(t)$ of the differential equation (14), it will be convenient to involve in the proof the trivial differential equation

$$r'/r = \alpha(t)\cos^2\theta + (\beta(t) + \gamma(t))\cos\theta\sin\theta + \delta(t)\sin^2\theta \tag{17}$$

for $r = r(t)$. The arrangement of the proof will make it possible to immediately extend the results concerning (3) and/or (7) to corresponding statements concerning certain non-linear, non-conservative binary systems.

In view of the remarks following (14), it can be supposed that the real parts of the characteristic numbers of (1) are negative. It is also clear that an affine transformation of the (x, y)-plane leaves the assumptions and assertions of (*) unchanged. Hence it can be supposed that (1) is in a normal form satisfying

$$a\cos^2 + \tfrac{1}{2}(b+c)\sin^2\theta + d\sin^2\theta \leqq -\text{const.} < 0 \text{ for all } \theta. \tag{18}$$

Accordingly, if $x = x(t)$, $y = y(t)$ is a non-trivial solution of (7), then $dr/dt < 0$, by (17), and (9) holds. Thus $r = r(t)$ has an inverse $t = t(r)$, and $t(r) \to \infty$, as $r \to 0$. Put

$$f(x, y) = (\alpha(t) - a)x + (\beta(t) - b)y, \quad g(x, y) = (\gamma(t) - c)x + (\delta(t) - d)y,$$

where $t = t(r) = t((x^2 + y^2)^{\frac{1}{2}})$. Then (5) holds by virtue of (8), since $t(r) \to \infty$ as $r \to 0$. Clearly, $x = x(t)$, $y = y(t)$ is a solution of (3). It follows from the results in [2], pp. 123-125, on the system (3)-(5) that $\theta(t)$ satisfies (11) in case (i), and (15) in case (ii), finally either (11) or (15) in case (iii).

It remains to be shown that the decision in the alternative of the case (iii) is independent of the choice of the solution. To this end, it can be supposed that (1) has the normal form in which

$$a = -1, \quad b = 0, \quad c = 1, \quad d = -1; \tag{19}$$

in particular, $\gamma(t) \neq 0$ for large t. Consider for large t the linear, second order differential equation

$$(p(t)z')' + h(t)z' + q(t)z = 0, \tag{20}$$

where

$$p(t) = \gamma^{-1}(t), \quad h(t) = \delta(t) - \alpha(t), \quad q(t) = -\beta(t). \tag{21}$$

If $z = z(t)$ is a non-trivial solution of (20), then a continuous determination of

$$\theta = \arctan\{z/(pz')\} \tag{22}$$

satisfies (14). Conversely, if $\theta = \theta(t)$ is a solution of (14), then a quadrature of (22) leads to a solution $z = z(t)$ of (20). (Actually, the quadrature in question is first just formal, because of the infinities of $\cot\theta(t)$, corresponding to the infinities of $\log|z(t)|$; so that the integration constants must be chosen on successive t-intervals, bounded by the points t of infinity of $\cot\theta(t)$, so as to insure that $z(t)$ has a continuous derivative.)

At a given t-value, $\theta(t)$ becomes 0 $(\mathrm{mod}\,\pi)$ if and only if $z(t) = 0$, and, at such a value of t, $\theta'(t) > 0$. Hence it is clear that (15) holds if and only if one and/or every non-trivial solution $z = z(t)$ of (20) has infinitely many zeros (as $t \to \infty$).

Consequently, in the normalized case (19) of (iii), the first or the second alternative of the cases (15) of (11), where $\theta_0 = 0 \ (\mathrm{mod}\,\pi)$, holds for every solution according as (20) is or is not oscillatory. This proves (*).

Remark. The argument just completed contains a proof of the following theorem (cf. [1], pp. 391-392): *If* (20) *is non-oscillatory for large* t *and*

$$p(t) \to 1, \qquad h(t) \to 0, \qquad q(t) \to 0 \text{ as } t \to \infty,$$

then, for every non-trivial solution $z = z(t)$ *of* (20),

$$z'(t)/z(t) \to 0 \text{ as } t \to \infty. \tag{23}$$

A careful perusal of the proof of (*) shows that (*) can be extended as follows:

(†) *Let $f(x, y, t)$, $g(x, y, t)$ be continuous functions for $x^2 + y^2 \leqq$ const., $0 \leqq t < \infty$ satisfying*

(5 bis) $\quad f/(x^2 + y^2)^{\frac{1}{2}} \to 0,\ g/(x^2 + y^2)^{\frac{1}{2}} \to 0$ as $(x, y, t) \to (0, 0, \infty)$.

Let the real parts of the characteristic numbers of the matrix (1) *be negative. Let (x_0, y_0) be sufficiently near, but not equal to, $(0, 0)$ and let $x = x(t)$, $y = y(t)$ be a solution (for small non-negative t) of the initial value problem*

(3 bis) $\quad x' = ax + by + f(x, y, t), \qquad y' = cx + dy + g(x, y, t)$

and $x(0) = x_0$, $y(0) = y_0$. Then the solution $x = x(t)$, $y = y(t)$ can be continued over the half-line $0 \leqq t < \infty$. The function $r = r(t)$ is positive and satisfies (9). *A continuous determination of* (10) *satisfies* (11) *when* (1) *is in the case* (i), *and* (15) *when* (1) *is in the case* (ii), *finally either* (11) *or* (15) *when* (1) *is in the case* (iii). *Furthermore, when* (11) *holds, $\theta = \theta_0$ is a zero of* (16); *in addition, $x'^2 + y'^2 \neq 0$ for a large t and a continuous determination of* (12) *satisfies* (13).

In contrast to the linear case of (*), the alternative in case (iii) can now depend on the solution; cf. [2], p. 123.

As to the first part of (†), that concerning the continuability of $(x(t), y(t))$ and (9), cf. [4], pp. 816-817.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "On non-oscillatory linear differential equations of second order," *American Journal of Mathematics*, vol. 74 (1952), pp. 389-400.

[2] ——— and A. Wintner. "On the behavior of the solutions of real binary differential systems at singular points," *ibid.*, vol. 75 (1953), pp. 117-126.

[3] O. Perron, "Über Stabilität und asymptotisches Verhalten der Integrale von Differentialgleichungssystemen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 129-160.

[4] A. Wintner, "Vortices and nodes," *American Journal of Mathematics*, vol. 69 (1947), pp. 815-824.

UMBILICAL POINTS AND W-SURFACES.*

By PHILIP HARTMAN and AUREL WINTNER.

A surface will be called a *special Weingarten surface* (special W-surface) if its Gaussian and mean curvatures, K and H, are connected by an identity

$$F(K,H)=0 \tag{1}$$

in which F satisfies the following conditions:

(*) *The function $F(K,H)$ is defined and of class C^2 on the portion $K\leqq H^2$ of the (K,H)-plane and satisfies*

$$\tfrac{1}{2}F_H+HF_K\neq 0 \quad \text{when } K=H^2. \tag{2}$$

Actually, the assumption that F is of class C^2 will be needed only on a set of points (K,H) on and near the parabola $K=H^2$. In terms of the principal curvatures k_1, k_2, the relation (1) can be written in the form

$$W(k_1,k_2)=0, \tag{3}$$

where $W(k_1,k_2)=F(k_1k_2,\tfrac{1}{2}(k_1+k_2))$. Correspondingly, (2) means that

$$\partial W/\partial k_1\cdot\partial W/\partial k_2>0 \quad \text{if} \quad k_1=k_2. \tag{4}$$

In fact, it is readily verified that the product of the partial derivatives in (4) is $(\tfrac{1}{2}F_H+HF_K)^2\geqq 0$ if $k_1=k_2$, and so the inequality in (2) implies that in (4).

This fact makes it clear that the above definition of a special W-surface is essentially that of Chern [1], who, without specifying the smoothness of W, requires that a relation of the type (3), (4) be satisfied. The assumption that $F(H,K)$ is of class C^1 or C^2 implies that (4) is of class C^1 or C^2, respectively. On the other hand, the assumption that $W(k_1,k_2)$ is smooth does not imply that $F(H,K)=W(H+(H^2-K)^{\frac{1}{2}},H-(H^2-K)^{\frac{1}{2}})$ is smooth at points where $K=H^2$. It is for this reason that a special W-surface was defined above in terms of a relation of the type (1), rather than one of the type (3). This definition also circumvents the hypothesis of H. Hopf [5], p. 234, that $\kappa=dk_1/dk_2$ exists at the points of the surface where $k_1=k_2$ (that is, $K=H^2$).

* Received November 23, 1953.

The main theorem to be proved is the following:

(I) *Let S be a (small piece of a) special W-surface of class C^2. Then, unless S is part of a plane or of a sphere, the umbilical points (if any) are isolated and their indices are negative.*

An umbilical point of S is a point where $H^2 = K$ ($\geqq 0$). According to this definition, a flat point (that is, a point where $H = K = 0$) is an umbilical point. It is understood that "index" in (I) refers to the net of the directions of principal curvature. The terminology (net, index) is the same as in [6], where self-contained proofs of what is needed are developed.

The use of elliptic partial differential equations in the proof of (I) will show that, by virtue of a theorem of Nirenberg [7], p. 106, the surface S has local parametrizations $z = z(x, y)$ in which the function $z = z(x, y)$ is of class C^3 (in fact, the third order partial derivatives of z satisfy uniform Hölder conditions) in a vicinity of the umbilical points.

The assertion (I) is contained in a theorem of Hopf ([5], p. 236) if the assumption that S is of class C^2 is strengthened to the assumption that S is analytic in a vicinity of its umbilical points. On the other hand, Hopf's proof is not applicable even if it is assumed that S is of class C^∞.

If a net of directions (say, principal directions of curvature) on a closed orientable surface of genus zero has only a finite number of singular points, then the index of at least one of those points is positive; cf. [6], p. 753. Hence, an immediate corollary of (I) is the following:

(II) *If a closed orientable surface S of genus 0 is a special W-surface of class C^2, then S is a sphere.*

This theorem was proved by Chern ([1], Theorem 1, p. 279) under the additional assumption that the Gaussian curvature K on S is positive throughout. Chern's smoothness assumption is that S be of class C^3; actually, a perusal of Chern's proof seems to indicate that his proof implicitly assumes that S is of class C^4; cf. his use of d^2k_1 on p. 283. Hopf ([5], p. 234) points out that a generalization of Chern's theorem can be proved along the lines of Hilbert's proof (cf., e. g., [8], pp. 122-123) that a closed surface of constant Gaussian curvature is a sphere. Since this proof requires that, near a non-umbilical point, S be of class C^2 in terms of a line of curvature parametrization, it would also seem to require implicitly that S be of class C^4; cf. the results of [4] for the possibly analogous situation in the problem of asymptotic parametrizations.

If S in (II) is assumed to be analytic, then (II) is contained in a theorem of Hopf [5], p. 237.

Proof of (1). Let S be represented as $z = z(x, y)$ for small $x^2 + y^2$. Then z is of class C^2 and satisfies the differential equation (1), in which

$$K = (rt - s^2)/(1 + p^2 + q^2)^2$$

$$H = ((1 + q^2)r - 2pqs + (1 + p^2)t)/2(1 + p^2 + q^2)^{\frac{3}{2}} \tag{5}$$

and $p = z_x$, $q = z_y$, $r = z_{xx}$, $s = z_{xy}$, $t = z_{yy}$. A simple calculation shows that if the left side of (1) is considered to be a function $\Phi(p, q, r, s, t)$ of p, q, r, s, t by virtue of (5), then $4\Phi_r\Phi_t - \Phi_s^2$, up to the factor $(1 + p^2 + q^2)^2$, is $F_H^2 + 4HF_HF_K + 4KF_K^2$, which is $4(\frac{1}{2}F_H + HF_K)^2$ when $K = H^2$. Hence, (2) means that the partial differential equation (1) is of elliptic type near an umbilical point. Correspondingly, a theorem of Nirenberg [7], p. 106, implies that if $(x, y) = (0, 0)$ is an umbilical point of S, then $z = z(x, y)$ is of class C^3 (and, in fact, its third order partial derivatives satisfy a uniform Hölder condition) in a vicinity of $(x, y) = (0, 0)$.

Flat points ($K = 0$). Suppose that

$$p(0, 0) = q(0, 0) = 0 \tag{6}$$

and that $(x, y) = (0, 0)$ is a flat point, that is, that $K = H = 0$ holds at $(x, y) = (0, 0)$. Then $(K, H) = (0, 0)$ satisfies (1). Hence, the plane $z \equiv z(0, 0)$ is a solution of the partial differential equations (1), (5). The results of [3], pp. 464-469, imply that if the two solutions $z = z(x, y)$, $z \equiv z(0, 0)$ are not identical, then the second order derivatives of the difference $z(x, y) - z(0, 0)$ satisfy asymptotic formulae of the form

$$\begin{aligned} z_{xx} &= \rho^n(c_1 \cos n\theta + c_2 \sin n\theta) + o(\rho^n), \\ z_{xy} &= \rho^n(c_2 \cos n\theta - c_1 \sin n\theta) + o(\rho^n), \qquad (\rho \to 0) \\ z_{yy} &= \rho^n(-c_1 \cos n\theta - c_2 \sin n\theta) + o(\rho^n), \end{aligned} \tag{7}$$

where n is a non-negative integer, $x = \rho \cos \theta$, $y = \rho \sin \theta$ and c_1, c_2 are constants for which $c_1^2 + c_2^2 \neq 0$. Since $(x, y) = (0, 0)$ is an umbilical point, the integer n is positive.

In particular, (7) implies that

$$z_{xx}z_{yy} - z_{xy}^2 = -\rho^{2n}(c_1^2 + c_2^2) + o(\rho^{2n}) \text{ as } \rho \to 0. \tag{8}$$

Consequently, the Gaussian curvature of S is negative for small positive $\rho^2 = x^2 + y^2$, and so $(x, y) = (0, 0)$ is an isolated umbilical point.

Since the net of the directions of principal curvature can be replaced by the net of asymptotic directions (the angles of which are bisected by the

directions of principal curvature), it follows that the index of $(0, 0)$ with respect to the net of the lines of curvature is the same as that with respect to the net of asymptotic lines, the net defined by

$$z_{xx}dx^2 + 2z_{xy}dxdy + z_{yy}dy^2 = 0. \tag{9}$$

Since the omission of the $o(\rho^n)$-terms of (7) in (9) is equivalent to a deformation of the net (9), it follows that the index of $(0, 0)$ with respect to the net of lines of curvature is $-n < 0$.

(Actually, the asymptotic formulae (7) are not needed to show that the index of $(0, 0)$ with respect to (9) is not positive. To this end, it is sufficient to know that $K(x, y) < 0$ for small $x^2 + y^2 > 0$; cf. [6], p. 756.)

Spherical points $(K > 0)$. Let (6) hold and suppose that $K > 0$. Then if

$$N = N(x, y) = (\lambda, \mu, \nu) \qquad (\lambda^2 + \mu^2 + \nu^2 = 1) \tag{10}$$

is a continuous unit normal to S, say $N(0, 0) = (0, 0, 1)$, then the spherical mapping $(x, y) \to (\lambda, \mu)$ is one-to-one. If z is of class C^2 or C^3, then the supporting function

$$w = w(\lambda, \mu) = x\lambda + y\mu + z\nu \tag{11}$$

is of class C^2 or C^3, respectively, in a vicinity of $(\lambda, \mu) = (0, 0)$; cf. [2], p. 310. In particular, the function $w(\lambda, \mu)$ is of class C^3 if $(x, y) = (0, 0)$ is a spherical point of S, that is, an umbilical point $(H^2 = K \geqq 0)$ which is not a flat point $(H = K = 0)$.

The function $F(1/V, U/2V)$ of $(U, V) = (2H/K, 1/K)$ is of class C^2 for $U^2 \geqq 4V > 0$. The equation (1) becomes a partial differential equation for $w(\lambda, \mu)$ by virtue of the Weingarten formulae

$$U = 2H/K = \Delta_2 w + 2w, \qquad V = 1/K = \Delta_{22} w + w\Delta_2 w + w^2,$$

where Δ_2 is the Laplace operator with respect to the third fundamental form,

$$|\, dN \,|^2 = \{(1 - \mu^2)d\lambda^2 + 2\lambda\mu d\lambda d\mu + (1 - \lambda^2)d\mu^2\}/\nu^2, \tag{12}$$

and Δ_{22} is the corresponding Monge-Ampère operator; cf., e. g., [2], p. 311.

If $(x, y) = (0, 0)$ is a spherical point of S, then the assumption (2) means that the partial differential equation (1) for w is of the elliptic type at, hence near, $(\lambda, \mu) = (0, 0)$; cf., e. g., [2], pp. 330-331.

If the Gaussian curvature of S at $(x, y) = (0, 0)$ is $K(0, 0) = 1/R^2$, then the sphere corresponding to the constant supporting function $w \equiv R$ is a solution of (1). Hence, the results of [3], pp. 464-469, imply that if

the origin on the z-axis is chosen so that $z(0,0) = R$ (hence $w(0,0 = R)$, then either $w(\lambda,\mu) \equiv R$ or the difference $w(\lambda,\mu) - R$ has second order partial derivatives which satisfy asymptotic formulae, analogous to (7), with $n \geqq 0$ (in fact, those which result by replacing x, y, z in (7) by λ, μ, w, respectively). Hence, quadratures show that the partial derivatives w_λ, w_μ are $O(\chi^{n+1})$ and that $w(\lambda,\mu)$ is $R + O(\chi^{n+2})$, where $\chi = (\lambda^2+\mu^2)^{\frac{1}{2}}$ and $\chi \to 0$.

Let $\alpha = \beta(\lambda,\mu)$, $\beta = \beta(\lambda,\mu)$ and $\gamma = \gamma(\lambda,\mu)$ denote the first, second and third fundamental matrices of S in its (λ,μ)-parametrization. Thus, γ is the matrix of the quadratic form (12) and

$$\gamma = \beta\alpha^{-1}\beta; \tag{13}$$

cf. [2], p. 304. If $\delta = \delta(\lambda,\mu)$ is the symmetric matrix the elements of which are the components of the second covariant derivative of w with respect to the quadratic form (12), then [1]

$$\beta = \delta + w\gamma. \tag{14}$$

The asymptotic estimates for the first partial derivatives of w show that $\delta(\lambda,\mu)$ is the Hessian matrix, say $\eta = \eta(\lambda,\mu)$, of w, up to an additive correction of $O(\chi^{n+1})$, as $\chi = (\lambda^2+\mu^2)^{\frac{1}{2}} \to 0$. Hence, the asymptotic formula for w shows that (14) can be written, as $\chi \to 0$, in the form

$$\beta - R\gamma = \eta + O(\chi^{n+1}). \tag{15}$$

Thus, by the asymptotic formula, analogous to (7), for the elements of η,

$$\det(\beta - R\gamma) = -\chi^{2n}(c_1^2 + c_2^2) + o(\chi^{2n}), \quad \text{as} \quad \chi \to 0.$$

In particular, $\det(\beta - R\gamma) < 0$ if $\chi > 0$ is sufficiently small. Hence, the characteristic numbers k_1, k_2 of $\beta^{-1}\gamma$ satisfy $k_1 < 1/R < k_2$ for small $\chi > 0$. Since $\beta^{-1}\gamma = \alpha^{-1}\beta$, by (13), the numbers k_1, k_2 are the principal curvatures. Consequently, these curvatures are distinct for small $\chi > 0$, so that $(\lambda,\mu) = (0,0)$ is an *isolated* umbilical point of S.

This consideration shows that the integer n is positive; for otherwise $(\lambda,\mu) = (0,0)$ were not an umbilical point of S.

Since (13) implies that $\beta - \gamma/k = \beta\alpha^{-1}(k\alpha - \beta)/k$ and since the directions $(d\lambda, d\mu)$ of principal curvature are the characteristic vectors of the (singular) matrices $k_1\alpha - \beta$, $k_2\alpha - \beta$, they are also the characteristic vectors

[1] In formulae (66) and (67) of [2], p. 312, the partial derivatives w_{ik} should be replaced by the covariant derivatives $\nabla_{ik}w$.

The discrepancy of a minus sign between the formula (14) of the above text and the formula (66) of [2] is due to a different orientation of the normal vector.

of $\beta - \gamma/k_1$, $\beta - \gamma/k_2$. In view of $k_1 < 1/R < k_2$, the net of directions of principal curvature can be replaced by the net of directions $(d\lambda, d\mu)$ for which the quadratic differential form belonging to $\beta - R\gamma$ vanishes (the bisectors of the angles of the latter net are the directions of principal curvature). It is seen from (15) that the latter net can be deformed into the net of directions $(d\lambda, d\mu)$ for which the quadratic differential form belonging to η vanishes, that is, into the net defined by

$$w_{\lambda\lambda}d\lambda^2 + 2w_{\lambda\mu}d\lambda d\mu + w_{\mu\mu}d\mu^2 = 0. \tag{16}$$

Hence, the index of $(\lambda, \mu) = (0, 0)$ relative to the net of directions of principal curvature is $-n < 0$. This proves (I).

Remark. Let (II*) denote that weakened form of (II) in which it is assumed that the Gaussian curvature on S is positive (as in Chern's theorem [1], p. 279). Then (II*) follows from the theorem (□) of the Appendix of [2], pp. 330-333.

Under more restrictive conditions of differentiability on F and S, the result (II*) is contained in a theorem of Pogoreloff (1948). Under the assumption that F and S are analytic, (II*) is implied by a result of Alexandroff (1939). It should be mentioned that, in the analytic case, the use of index theory in differential geometry is due to Cohn-Vossen (1927) and the combination of indices with the theory of elliptic partial differential equations to H. Lewy (1938); cf. also Hopf and Samelson (1938), Alexandroff (1939) and Zhitomirsky (1939). For references to the papers mentioned, see [2], p. 334 and [3], p. 476.

Remarks on the Appendix in [2]. In theorem (□) of [2], p. 330 (in which the last word should be *satisfy* instead of *satifies*), it was assumed that the surface S is of class C^3. Since the function F occurring in the elliptic partial differential equation $F = 0$ is assumed *loc. cit.* to be of class C^2, it follows from Nirenberg's above-quoted result ([7], p. 106), which has appeared in the meantime, that it is sufficient to assume only that S is of class C^2, for S will then be of class C^3 by necessity.

Another improvement on the Appendix of [2] is as follows: On pp. 332-333, the asymptotic formula for $\det(w_{ik} - \omega_{ik})$ was used to show that the congruence points are isolated. The proof that the index of a congruence point is negative (is omitted but) follows at once from the asymptotic formulae for $w_{ik} - \omega_{ik}$ (analogous to (7) above) supplied by the result of

[3]. (That the index is just non-positive, instead of being negative, can also be concluded from a theorem of Hopf and Samelson, [6], p. 756, top.)

The paper referred to by [3] on p. 334 of [2] is the paper [3] in the bibliography below.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] S. S. Chern, "Some new characterizations of the euclidean sphere," *Duke Mathematical Journal*, vol. 12 (1945), pp. 279-290.

[2] P. Hartman and A. Wintner, "On the third fundamental form of a surface," *American Journal of Mathematics*, vol. 75 (1953), pp. 298-334.

[3] ——— and A. Wintner, "On the local behavior of solutions of non-parabolic partial differential equations," *ibid.*, vol. 75 (1953), pp. 449-476.

[4] ——— and A. Wintner, "On asymptotic parametrizations," *ibid.*, vol. 75 (1953), pp. 488-496.

[5] H. Hopf, "Über Flächen mit einer Relation zwischen den Hauptkrümmungen," *Mathematische Nachrichten*, vol. 4 (1950-1951), pp. 232-249.

[6] ——— and H. Samelson, "Zur Beweis des Kongruenzsatzes für Eiflächen," *Mathematische Zeitschrift*, vol. 43 (1938), pp. 749-766.

[7] L. Nirenberg, "On nonlinear partial differential equations and Hölder continuity," *Communications on Pure and Applied Mathematics*, vol. 6 (1953), pp. 103-156.

[8] D. J. Struik, *Lectures in classical differential geometry*, 1950.

SOME FUNDAMENTAL FORMULAS FOR HERMITIAN MANIFOLDS WITH NON-VANISHING TORSION.*

By WILLIAM M. BOOTHBY.

---

1. Introduction. In this paper is developed for the case of complex-analytic manifolds with a Hermitian metric the useful technique originated by Chern which consists in referring to the principal fibre bundle B over a manifold M the tensor calculus of M. The first part of the paper is an exposition of general facts about complex manifolds which are already well known, including a brief discussion of the method of Chern in the complex-analytic case. In sections 10-13 some basic formulas are worked out including those for $d\theta$, $*\theta$, $\delta\theta$ for a differential form θ of M referred to B. These are generalizations of the formulas given by Guggenheimer for the case of a Hermitian-Kähler metric. The final paragraphs give applications of these formulas which use Bochner's Lemma and which are similar to the applications given by Bochner in the case of a Hermitian-Kähler metric.

This study was undertaken at the suggestion of S. S. Chern to whom the author is deeply grateful for helpful comments and discussions during its progress.

2. Complex manifolds. Let M be a topological manifold of dimension $2n$ and let $\{U\}$ be a family of open sets covering M in each of which is given a system of complex coordinates $u^1, \cdots, u^n$. Further assume that wherever two coordinate neighborhoods U, V overlap the corresponding change of coordinates is given by complex-analytic functions $v^i = v^i(u^1, \cdots, u^n)$ with non-zero Jacobian. The family of coordinate neighborhoods is then said to define a complex-analytic structure on M. For functions defined in some neighborhood of $p \in M$, not necessarily the same for each function, the concept of being analytic at p is then invariant and we let $\mathcal{A}(p)$ denote the family of all functions which are analytic at p. The complex-analytic structure automatically defines a real-analytic structure: corresponding to complex coordinates $u^1, \cdots, u^n$ we have real coordinates $x^1, \cdots, x^{2n}$ where [1]

* Received October 9, 1953.

[1] We adopt the following summation conventions: latin indices i, j, k, etc., run from 1 to n; greek indices α, β, γ, etc. from 1 to $2n$. For indices from 1 to n a star (*)

$u^j = x^j + ix^{j^*}$. The change from real coordinates x^α to real coordinates y^α corresponding to $v^j = y^j + iy^{j^*}$ is thus necessarily given by real-analytic functions with positive Jacobian. Any real-analytic function may be written as a formal power series in $u^1, \cdots, u^n, u^{*1}, \cdots, u^{*n}$ by replacing x^j by $\frac{1}{2}(u^j + u^{*j})$ and x^{j^*} by $\frac{1}{2i}(u^j - u^{*j})$, so we may and shall use $u^1, \cdots, u^n$, $u^{*1}, \cdots, u^{*n}$ in lieu of $x^1, \cdots, x^{2n}$ wherever real-analytic coordinates are required. We let $\mathfrak{F}(p)$ denote the family of all complex-valued functions $f = u + iv$ whose real and imaginary parts are real-analytic, i. e., the family of all functions having expansions in powers of $u^1, \cdots, u^n, u^{*1}, \cdots, u^{*n}$, local coordinates at p. Then at p both $\partial f/\partial u^j$ and $\partial f/\partial u^{*j}$ are defined for $f \in \mathfrak{F}(p)$, and the subset $\mathfrak{A}(p)$ of $\mathfrak{F}(p)$ consists of exactly those functions f for which $\partial f/\partial u^{*j} = 0$, $j = 1, \cdots, n$. We let $\mathfrak{A}^*(p)$ denote the functions whose complex conjugates are analytic at p, i. e. the functions f for which $\partial f/\partial u^j = 0$ for $j = 1, \cdots, n$. $\mathfrak{F}(p)$ is a linear space over the complex numbers and $\mathfrak{A}(p)$, $\mathfrak{A}^*(p)$ are subspaces. Complex conjugation, $f \to f^*$ is a conjugate automorphism of $\mathfrak{F}(p)$ carrying $\mathfrak{A}(p)$ onto $\mathfrak{A}^*(p)$ and conversely.

3. The vectors tangent to M. The definition of a tangent vector at $p \in M$ will be the complex analog of the invariant definition given by Chevalley [4] for real-analytic manifolds. A vector at p is any linear map $\boldsymbol{v}$ of $\mathfrak{F}(p)$ into the complex numbers which on a composite function $f = f(q^1, \cdots, q^r)$, $q^1, \cdots, q^r$, $f \in \mathfrak{F}$, satisfies the chain rule

$$\boldsymbol{v}(f) = \boldsymbol{v}(q^1)\partial f/\partial q^1 + \cdots + \boldsymbol{v}(q^r)\partial f/\partial q^r, .$$

from which it follows easily that the collection of all such vectors forms a vector space of dimension $2n$, $W^{2n}(p)$ over the complex numbers. If $u^1, \cdots, u^n$ are coordinates at p, then every $f \in \mathfrak{F}$ is a composite function $f = f(u^1, \cdots, u^n, u^{*1}, \cdots, u^{*n})$ and this gives rise to a natural basis $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n, \boldsymbol{f}^*_1, \cdots, \boldsymbol{f}^*_n$ of W^{2n} associated with the coordinates and characterized by $\boldsymbol{f}_\alpha(u^\beta) = \delta_{\alpha\beta}$. Although we do not do so, it is suggestive to use the notation $\boldsymbol{f}_j = \partial/\partial u^j$ and $\boldsymbol{f}^*_j = \partial/\partial u^{*j}$, since if $\boldsymbol{v} \in W^{2n}$ and $f = f(u^1, \cdots, u^n, u^{*1}, \cdots, u^{*n}) \in \mathfrak{F}$, we have

$$\boldsymbol{v}(f) = \boldsymbol{v}(u^\alpha)\partial f/\partial u^\alpha = \boldsymbol{v}(u^\alpha)\boldsymbol{f}_\alpha(f).$$

---

affixed to the index automatically adds n, and for those from $n+1$ to $2n$ subtracts n. Complex conjugates are denoted by a star (*) rather than a bar ($^-$) and for symbols representing complex numbers the placing of a star on the index is equivalent to taking the complex conjugate. For example we write u^{*j} or u^{j^*} for the conjugate of u^j. (This notation is adopted for typographical convenience. A better notation is to replace all stars, including those on indices, by bars, see [6]). Repeated indices are to be summed.

We remark that since the coordinates u^j are valid in a neighborhood of p, they furnish a natural basis for the tangent space $W^{2n}(q)$ at each point q in the neighborhood, thus $\boldsymbol{f}_\alpha = \partial/\partial u^\alpha$ actually is a field of bases over the coordinate neighborhood. If the coordinate neighborhood V overlaps with U, we have $\partial/\partial v^\alpha = (\partial u^\beta/\partial v^\alpha)(\partial/\partial u^\beta)$, that is, $\boldsymbol{g}_\alpha = (\partial u^\beta/\partial v^\alpha)\boldsymbol{f}_\beta$.

The interesting point is that W^{2n} carries further structure in the form of certain maps and subspaces which may be invariantly defined in our schema. First we may define a conjugate automorphism $\boldsymbol{v} \to \boldsymbol{v}^*$ of W^{2n} by

$$\boldsymbol{v}(f) = [\boldsymbol{v}(f^*)]^*, \qquad f \in \mathfrak{F}.$$

We shall refer to this map as the *conjugate automorphism* of W^{2n}. For the natural basis $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n, \boldsymbol{f}^*_1, \cdots, \boldsymbol{f}^*_n$ defined just above we have $\boldsymbol{f}^*_j = \boldsymbol{f}_{j^*}$. Second we may define subspaces V^n (and V^{*n}) of W^{2n} as those vectors of W^{2n} whose value on $\mathfrak{A}^*(p)$ (and $\mathfrak{A}(p)$ respectively) is zero. Then we may prove the following statement.

PROPOSITION 1. *W^{2n} is the direct sum of its subspaces V^n and V^{*n}, each of which is of dimension n. The conjugate automorphism carries V^n onto V^{*n} and conversely. In any natural basis relative to local coordinates $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n \in V^n$ and $\boldsymbol{f}^*_1, \cdots, \boldsymbol{f}^*_n \in V^n$.*[2]

Proof. It is clear from the definitions that the conjugate automorphism carries V^n onto V^{*n} and V^{*n} onto V^n, so that both subspaces have the same dimension. If any vector $\boldsymbol{v}$ is in both V^n and V^{*n}, then $\boldsymbol{v}(u^j) = 0 = \boldsymbol{v}(u^{*j})$, $j = 1, \cdots, n$ for local coordinates $u^1, \cdots, u^n$ at p; and therefore by the chain rule $\boldsymbol{v}(f) = 0$ for every $f \in \mathfrak{F}$, so that $\boldsymbol{v} = 0$. It remains to show only that $V^n + V^{*n}$ spans W^{2n}; but it is easily seen from their definitions that $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n \in V^n$ and $\boldsymbol{f}^*_1, \cdots, \boldsymbol{f}^*_n \in V^{*n}$ and these vectors form a basis of W^{2n}. We shall call the vectors of V^n *pure complex* vectors, since, in the sense that were we to restrict ourselves entirely to complex-analytic functions in defining the differentiable structure on M, M being then a manifold of complex dimension n, we could proceed by analogy with an n-dimensional real-analytic manifold, and define just the n-dimensional vector space V^n over the complex numbers as tangent vectors to M. However, for most purposes restriction to complex-analytic functions is too severe, and we are impelled to consider real-analytic functions. It is in this latter context that it seems natural, though not essential, to introduce the vector space W^{2n}.

A vector will be called *real* if its restriction, as a mapping, to the real-valued functions of $\mathfrak{F}(p)$ is real-valued, that is if its value on every real-

[2] It is convenient to denote vectors of V^{*n} by $\boldsymbol{v}^*$, i. e., image of $\boldsymbol{v} \in V^n$.

valued function in $\mathfrak{F}(p)$ is a real number. The real vectors in W^{2n} form a subset R^{2n} which is closed under addition and multiplication by real scalars and is therefore a vector space over the real numbers.

PROPOSITION 2. *A vector $\boldsymbol{v}$ of W^{2n} is in R^{2n} if and only if $\boldsymbol{v}^* = \boldsymbol{v}$; the dimension of R^{2n} is indeed $2n$; and each vector $\boldsymbol{v}$ of W^{2n} may be uniquely written in the form $\boldsymbol{v} = \boldsymbol{v}_1 + i\boldsymbol{v}_2$ where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are real.*

Proof. Let $\boldsymbol{v} \in R^{2n}$ and $f \in \mathfrak{F}(p)$, then $f = f_1 + if_2$ where $f_1, f_2 \in \mathfrak{F}(p)$ and are real so that $\boldsymbol{v}(f_1)$, $\boldsymbol{v}(f_2)$ are real numbers. Since $\boldsymbol{v}$ is linear, $\boldsymbol{v}(f) = \boldsymbol{v}(f_1) + i\boldsymbol{v}(f_2)$. On the other hand

$$\boldsymbol{v}(f) = [\boldsymbol{v}(f^*)]^* = [\boldsymbol{v}(f_1) - i\boldsymbol{v}(f_2)]^* = \boldsymbol{v}(f_1) + i\boldsymbol{v}(f_2).$$

Then $\boldsymbol{v}^* = \boldsymbol{v}$. Suppose on the other hand that $\boldsymbol{v}^* = \boldsymbol{v}$ and that $f \in \mathfrak{F}(p)$ is a real-valued function. Then $[\boldsymbol{v}(f)]^* = \boldsymbol{v}^*(f^*) = \boldsymbol{v}^*(f) = \boldsymbol{v}(f)$, that is the value of $\boldsymbol{v}$ on f is a real number, whence $\boldsymbol{v} \in R^{2n}$. Now the dimension of R^{2n} is $\geqq 2n$ since the $2n$ vectors

$$\tfrac{1}{2}(\boldsymbol{f}_1 + \boldsymbol{f}^*_1), \cdots, \tfrac{1}{2}(\boldsymbol{f}_n + \boldsymbol{f}^*_n), \qquad -\tfrac{1}{2}i(\boldsymbol{f}_1 - \boldsymbol{f}^*_1), \cdots, -\tfrac{1}{2}i(\boldsymbol{f}_n - \boldsymbol{f}^*_n)$$

are in R^{2n} and form a basis of W^{2n} and must therefore be independent over the real numbers. On the other hand if $\alpha^1\boldsymbol{v}_1 + \cdots + \alpha^r\boldsymbol{v}_r = 0$ is a combination of real vectors $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r$ with complex numbers as coefficients, then, since $\boldsymbol{v}_j = \boldsymbol{v}^*_j$, the conjugate automorphism applied to both sides gives $\alpha^{*1}\boldsymbol{v}_1 + \cdots + \alpha^{*r}\boldsymbol{v}_r = 0$ so that the same combination with only the real parts or with only the imaginary parts of the α^j's is zero. From this and the fact that every set of more than $2n$ vectors of W^{2n} is dependent over the complex numbers, we see that the dimension of $R^{2n} \leqq 2n$. Hence it is exactly $2n$ and the vectors above form a basis. It is worth remarking that the above basis of R^{2n} corresponds exactly to the natural basis $\partial/\partial x^1, \cdots, \partial/\partial x^n$, $\partial/\partial x^{1^*}, \cdots, \partial/\partial x^{n^*}$ corresponding to the real-analytic coordinates $x^1, \cdots, x^n$, $x^{1^*}, \cdots, x^{n^*}$ mentioned in section 2, and

$$\partial/\partial w^j = \partial/\partial x^j + i\partial/\partial x^{j^*}, \qquad \partial/\partial w^{*j} = \partial/\partial x^j - i\partial/\partial x^{j^*}.$$

Finally let $\boldsymbol{v}$ be an arbitrary vector in W^{2n}. It is easy to see that $\boldsymbol{v} = \frac{1}{2}(\boldsymbol{v} + \boldsymbol{v}^*) + \frac{1}{2i}(\boldsymbol{v} - \boldsymbol{v}^*)$ is a unique representation of $\boldsymbol{v}$ as a combination $\boldsymbol{v}_1 + i\boldsymbol{v}_2$ of real vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$.

Finally we define the automorphism Ψ (see Eckmann and Guggenheimer [5] and Guggenheimer [6]) introduced by Hopf. If $\boldsymbol{v} \in V^n$, then let $\Psi\boldsymbol{v} = i\boldsymbol{v}$ and, if $\boldsymbol{v} \in V^{*n}$, let $\Psi\boldsymbol{v} = -i\boldsymbol{v}$. Since $W^{2n} = V^n + V^{*n}$ this defines Ψ for an arbitrary vector of W^{2n}. The following statements about Ψ are clear.

PROPOSITION 3. *Ψ commutes with the conjugate automorphism: $\Psi \boldsymbol{v}^* = [\Psi \boldsymbol{v}]^*$ and V^n, V^{*n} are the invariant subspaces associated with the two proper values of $\pm i$ of Ψ.*

4. Inner products. Let $\boldsymbol{v} \cdot \boldsymbol{w}$ be a symmetric, bilinear, complex-valued function on W^{2n}, so that

(1) $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v}$ for $\boldsymbol{v}, \boldsymbol{w} \in W^{2n}$

(2) $[\alpha \boldsymbol{v} + \beta \boldsymbol{v}'] \cdot \boldsymbol{w} = \alpha[\boldsymbol{v} \cdot \boldsymbol{w}] + \beta[\boldsymbol{v}' \cdot \boldsymbol{w}]$
$\boldsymbol{v}[\alpha \boldsymbol{w} + \beta \boldsymbol{w}'] = \alpha[\boldsymbol{v} \cdot \boldsymbol{w}] + \beta[\boldsymbol{v} \cdot \boldsymbol{w}']$ for α, β complex numbers,

and further satisfying

(3) $[\boldsymbol{v} \cdot \boldsymbol{w}]^* = \boldsymbol{v}^* \cdot \boldsymbol{w}^*$

(4) $\boldsymbol{v} \cdot \boldsymbol{v} \geqq 0$, with equality if and only if $\boldsymbol{v} = 0$, and

(5) $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ if $\boldsymbol{v}$ and $\boldsymbol{w}$ are both in V^n or both in V^{*n}.

Then this function defines a real-valued inner product on R^{2n}. Moreover, the function $(\boldsymbol{v}, \boldsymbol{w})$ defined for vectors of V^n by

$$(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{w}^* \tag{6}$$

is a Herimitian scalar product on V^n. Conversely a Hermitian scalar product on V^n allows us to define a function with properties (1)-(5) on W^{2n}, namely

$$[\boldsymbol{v} + \boldsymbol{v}^{*\prime}] \cdot [\boldsymbol{w} + \boldsymbol{w}^{*\prime}] = (\boldsymbol{v}, \boldsymbol{w}') + (\boldsymbol{w}, \boldsymbol{v}') \tag{7}$$

where $\boldsymbol{v} + \boldsymbol{v}^{*\prime}$ and $\boldsymbol{w} + \boldsymbol{w}^{*\prime}$ are the unique representations of (arbitrary) vectors of W^{2n} as the sum of a vector of V^n and a vector of V^{*n}. Let $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n, \boldsymbol{f}^*_1, \cdots, \boldsymbol{f}^*_n$ be a basis of W^{2n} and let $g_{\alpha\beta} = \boldsymbol{f}_\alpha \cdot \boldsymbol{f}_\beta$. Then we have

$$g_{ij} = 0 = g_{i^*j^*} \text{ and } g_{ij^*} = g_{j^*i} = g^*_{ji^*} = g^*_{i^*j}. \tag{8}$$

If at each point p of M a function $\boldsymbol{v} \cdot \boldsymbol{w}$ as described is given on W^{2n} such that as functions of local coordinates relative to the natural basis the $g_{\alpha\beta}$ are real-analytic, then M is called a *Hermitian* manifold. Clearly it is also a Riemannian manifold. We remark merely that relative to such an inner product the map Ψ of section 3 is a rotation on R^{2n}, so that

$$\Psi \boldsymbol{v} \cdot \Psi \boldsymbol{w} = \boldsymbol{v} \cdot \boldsymbol{w}. \tag{9}$$

Relative to such a product on W^{2n} is defined a natural isomorphism from W^{2n} to W'^{2n} with $\boldsymbol{w}'$, the image of $\boldsymbol{w}$, defined by

$$\langle \boldsymbol{v}, \boldsymbol{w}' \rangle = \boldsymbol{v} \cdot \boldsymbol{w} \text{ for } \boldsymbol{v} \in W^{2n}. \tag{10}$$

Let $\phi^1, \cdots, \phi^n, \phi^{*1}, \cdots, \phi^{*n}$ be dual to $f_1, \cdots, f_n, f^*_1, \cdots, f^*_n$ a basis of W^{2n} and let $\xi^\alpha f_\alpha$ be a vector of W^{2n} with components $\xi^1, \cdots, \xi^{2n}$. If $\mu_\beta \phi^\beta$ is the image in W'^{2n} of $\xi^\alpha f_\alpha$ we have

$$\mu_i = \langle f_i, \mu_\beta \phi^\beta \rangle = f_i \cdot \xi^\alpha f_\alpha = \xi^{j^*} g_{ij^*} \text{ and,}$$

$$\mu_{i^*} = \langle f^*_i, \mu_\beta \phi^\beta \rangle = f^*_i \cdot \xi^\alpha f_\alpha = \xi^j g_{i^* j}. \tag{11}$$

In tensor notation these are standard formulas for lowering indices. By virtue of (8) they may be combined into

$$\mu_\alpha = g_{\alpha\beta} \xi^\beta. \tag{12}$$

5. Tensors on M. A tensor T at p of type (r, s) is an element of the Kronecker product $W^r{}_s = W^{2n} \otimes \cdots \otimes W^{2n} \otimes W'^{2n} \otimes \cdots \otimes W'^{2n}$ of r copies of W^{2n} and s copies of W'^{2n}. If a tensor of type (r, s) is given at each point of M, we have a tensor field $T(p)$ on M. Let $u^1, \cdots, u^n$ be local coordinates with $f_1, \cdots, f_n, f^*_1, \cdots, f^*_n$ and $\phi^1, \cdots, \phi^n, \phi^{*1}, \cdots, \phi^{*n}$ the corresponding dual bases in W^{2n} and W'^{2n}. Then we have

$$T(p) = T^{\beta_1, \cdots, \beta_r}{}_{\alpha_1, \cdots, \alpha_s} f_{\beta_1} \otimes \cdots \otimes f_{\beta_r} \otimes \phi^{\alpha_1} \otimes \cdots \otimes \phi^{\alpha_s}.$$

The components $T^{\beta_1, \cdots, \beta_r}{}_{\alpha_1, \cdots, \alpha_s}$ are in our case assumed to be real-analytic functions of the local coordinates $u^1, \cdots, u^n$ and their conjugates. Corresponding to the subspaces V^n and V^{*n} of W^{2n} and their duals, we have a direct sum decomposition of $W^r{}_s$ into subspaces. For example $W^3{}_0 = W^{2n} \otimes W^{2n} \otimes W^{2n}$ is the direct sum of 2^3 subspaces such as $V^{*n} \otimes V^{*n} \otimes V^n$, $V^n \otimes V^{*n} \otimes V^n$, etc. Tensors in these subspaces have, relative to a natural basis, all components other than those of the form $T_{i^* j^* k}$, $T_{ij^* k}$, etc. respectively, equal to zero. The elements of these subspaces are called *pure* tensors, and in studying linear operations on tensors one may conveniently restrict oneself to pure tensors, as will be done below. Tensor components written with latin indices may always be assumed the components of a pure tensor.

An important subset of $W^r{}_s$ consists of linear combinations with real coefficients of elements of $W^r{}_s$ of the form $u_1 \otimes \cdots \otimes u_r \otimes \theta^1 \otimes \cdots \otimes \theta^s$ where u_i are elements of $R^{2n} \subset W^{2n}$ and θ^i elements of $R'^{2n} \subset W'^{2n}$. This subset is naturally isomorphic to the Kronecker product

$$R^r{}_s = R^{2n} \otimes \cdots \otimes R^{2n} \otimes R'^{2n} \otimes \cdots \otimes R'^{2n}$$

over the real numbers of r copies of R^{2n} and s copies of R'^{2n} which will be thought of as thus imbedded in $W^r{}_s$. Bochner [1] has given the simple

condition which follows, that a tensor $\boldsymbol{T}$ of $W^r{}_s$ lie in $R^r{}_s$, i. e. be a real tensor. For definiteness consider again the space $W^3{}_0 = W^{2n} \bigotimes W^{2n} \bigotimes W^{2n}$ and let $\boldsymbol{T} = T^{\alpha\beta\gamma}\boldsymbol{v}_\alpha \bigotimes \boldsymbol{v}_\beta \bigotimes \boldsymbol{v}_\gamma$ be an element thereof. We define on $W^3{}_0$ the conjugate automorphism $\boldsymbol{T} \to \boldsymbol{T}^*$ by

$$\boldsymbol{T}^* = T^{*\alpha\beta\gamma}\boldsymbol{v}^*{}_\alpha \bigotimes \boldsymbol{v}^*{}_\beta \bigotimes \boldsymbol{v}^*{}_\gamma.$$

Now every vector $\boldsymbol{v}_\alpha$ has a unique decomposition $\boldsymbol{v}_\alpha = \boldsymbol{u}_\alpha + i\boldsymbol{w}_\alpha$, $\boldsymbol{u}_\alpha$ and $\boldsymbol{w}_\alpha$ real vectors, and also $T^{\alpha\beta\gamma} = R^{\alpha\beta\gamma} + iS^{\alpha\beta\gamma}$ with $R^{\alpha\beta\gamma}$ and $S^{\alpha\beta\gamma}$ real numbers. Thus $\boldsymbol{T}$ may be written

$$\boldsymbol{T}^{\alpha\beta\gamma} = (R^{\alpha\beta\gamma} + iS^{\alpha\beta\gamma})\,(\boldsymbol{u}_\alpha + i\boldsymbol{w}_\alpha) \bigotimes (\boldsymbol{u}_\beta + i\boldsymbol{w}_\beta) \bigotimes (\boldsymbol{u}_\gamma + i\boldsymbol{w}_\gamma).$$

Using the linearity in each of the " factors " it is seen that

$$\boldsymbol{T} = \boldsymbol{P} + i\boldsymbol{Q}, \text{ where } \boldsymbol{P}, \boldsymbol{Q} \,\varepsilon\, R^{2n} \bigotimes R^{2n} \bigotimes R^{2n}$$

to which $\boldsymbol{T}$ itself belongs then if and only if $\boldsymbol{Q} = 0$. In this notation $\boldsymbol{T}^*$ becomes

$$\boldsymbol{T}^* = \boldsymbol{P} - i\boldsymbol{Q},$$

so that $\boldsymbol{T} = \boldsymbol{T}^*$ is the condition for $\boldsymbol{T}$ to be in $R^{2n} \bigotimes R^{2n} \bigotimes R^{2n}$, i. e. to be real. $\boldsymbol{T}$ is said by Bochner to be *self-adjoint* in this case.

If an inner product is defined on W^{2n} we may use the corresponding natural isomorphism of W^{2n} to W'^{2n} to raise and lower indices of the components of tensors. Thus relative to dual bases $T^{\beta_1,\ldots,\beta_r}{}_{\alpha_1,\ldots,\alpha_s}$ is carried to, e. g. $T^{\beta_1\ldots\beta_r\gamma_1\ldots\gamma_s}$ where

$$T^{\beta_1\ldots\beta_r\gamma_1\ldots\gamma_s} = T^{\beta_1\ldots\beta_r}{}_{\alpha_1\ldots\alpha_s} g^{\alpha_1\gamma_1} g^{\alpha_2\gamma_2} \cdots g^{\alpha_s\gamma_s}.$$

If the basis chosen in W^{2n} is unit orthogonal in V^n, then $g_{ij^*} = \delta_{ij}$ and

$$T^{\beta_1\ldots\beta_r\gamma^*{}_1\ldots\gamma^*{}_s} = T^{\beta_1\ldots\beta_r}{}_{\gamma_1\ldots\gamma_s}.$$

We remark that using multi-linearity of elements of a Kronecker product it is easy to derive the classical formulas for change of components of a tensor under change of basis. If $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n, \boldsymbol{f}^*{}_1, \cdots, \boldsymbol{f}^*{}_n$ and $\boldsymbol{g}_1, \cdots, \boldsymbol{g}_n, \boldsymbol{g}^*{}_1, \cdots, \boldsymbol{g}^*{}_n$ are bases of W^{2n} related by $\boldsymbol{f}_i = P^j{}_i \boldsymbol{g}_j$, $\boldsymbol{f}^*{}_i = \boldsymbol{P}^{*j}{}_i \boldsymbol{g}^*{}_j$ and $\boldsymbol{T}$ is, for example, a tensor of type $(3, 0)$

$$\boldsymbol{T} = T^{\alpha\beta\gamma}\boldsymbol{f}_\alpha \bigotimes \boldsymbol{f}_\beta \bigotimes \boldsymbol{f}_\gamma$$

then we have

$$\boldsymbol{T} = T^{\alpha\beta\gamma}P^\lambda{}_\alpha \boldsymbol{g}_\lambda \bigotimes P^\mu{}_\beta \boldsymbol{g}_\mu \bigotimes P^\nu{}_\gamma \boldsymbol{g}_\nu = T^{\alpha\beta\gamma}P^\lambda{}_\alpha P^\mu{}_\beta P^\nu{}_\gamma \boldsymbol{g}_\lambda \bigotimes \boldsymbol{g}_\mu \bigotimes \boldsymbol{g}_\nu$$

so that the components of $\boldsymbol{T}$ relative to $\boldsymbol{g}_\alpha$ are $T^{\alpha\beta\gamma}P^\lambda{}_\alpha P^\mu{}_\beta P^\nu{}_\gamma$. This is the usual formula in the case of change of coordinates from u to v with

$P^{\lambda}{}_{\mu} = \partial v^{\lambda}/\partial u^{\mu}$. This formula has been used classically to define a tensor and for brevity we shall revert to the classical definitions for tensorial differential forms. For example a tensorial differential form of order p and type $(3, 0)$ is given if in each local coordinate system $u^1, \cdots, u^n$ are given $(2n)^3$ differential forms $\Omega^{\alpha\beta\gamma}$ of order p in du^i, du^{*i} transforming according to the rule given above for the change of components of $\boldsymbol{T}$. A tensor of type (r, s) is a tensorial form of order 0 and type (r, s).

6. Affine connections. There are two methods of computing a metric affine connection on a Hermitian manifold M, that is to determine for each local coordinate system $u^1, \cdots, u^n$ a set of $(2n)^2$ linear differential forms $\theta^{\beta}{}_{\alpha} = \Gamma^{\beta}{}_{\alpha\nu} du^{\nu}$ defining by $d\boldsymbol{f}_{\alpha} = \theta^{\beta}{}_{\alpha}\boldsymbol{f}_{\beta}$ the differentials $d\boldsymbol{f}_{\alpha}$ of the natural basis vectors $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n, \boldsymbol{f}_{1^*}, \cdots, \boldsymbol{f}_{n^*}$ in a manner independent of local coordinates. This latter means that if $v^1, \cdots, v^n$ is a coordinate system overlapping $u^1, \cdots, u^n$ and $\boldsymbol{g}_1, \cdots, \boldsymbol{g}_n, \boldsymbol{g}_{1^*}, \cdots, \boldsymbol{g}_{n^*}$ its corresponding basis (where we have [3] then $\boldsymbol{f}_{\alpha} = P^{\beta}{}_{\alpha}\boldsymbol{g}_{\beta}$ with $P^{\beta}{}_{\alpha} = \partial v^{\beta}/\partial u^{\alpha}$ and $\boldsymbol{g}_{\alpha} = Q^{\beta}{}_{\alpha}\boldsymbol{f}_{\beta}$ with $Q^{\beta}{}_{\alpha} = \partial u^{\beta}/\partial v^{\alpha}$), then we may compute $d\boldsymbol{f}_{\alpha}$ using either $\theta^{\beta}{}_{\alpha}$ of the u-coordinates or $\gamma^{\beta}{}_{\alpha}$ of the v-coordinates, and both must give the same result:

$$(1) \qquad d\boldsymbol{f}_{\alpha} = d(P^{\beta}{}_{\alpha}\boldsymbol{g}_{\beta}) = (dP^{\beta}{}_{\alpha} + P^{\nu}{}_{\alpha}\gamma^{\beta}{}_{\nu})\boldsymbol{g}_{\beta}, \text{ where } d\boldsymbol{g}_{\alpha} = \gamma^{\beta}{}_{\alpha}\boldsymbol{g}_{\beta}$$

whence we must have

$$(2) \qquad \theta^{\beta}{}_{\alpha} = (dP^{\mu}{}_{\alpha} + P^{\nu}{}_{\alpha}\theta^{\mu}{}_{\nu})Q^{\beta}{}_{\mu}.$$

This is the classical formula for the change of components of an affine connection with change of coordinates. We impose the condition

$$(3) \qquad d\boldsymbol{f}^*{}_i = [d\boldsymbol{f}_i]^* \quad \text{(i. e. } \theta^{*\beta}{}_{\alpha} = \theta^{\beta^*}{}_{\alpha^*}\text{)}$$

as part of the definition of affine connection on a complex manifold. The connection is called *metric* if

$$(4) \qquad d[\boldsymbol{f}_{\alpha}\cdot\boldsymbol{f}_{\beta}] = d\boldsymbol{f}_{\alpha}\cdot\boldsymbol{f}_{\beta} + \boldsymbol{f}_{\alpha}\cdot d\boldsymbol{f}_{\beta}.$$

If we define a *real* displacement dp of $p \in M$ by $dp = du^i\boldsymbol{f}_j + du^{*j}\boldsymbol{f}^*{}_j$ and a *pure complex* displacement by $dp = du^j\boldsymbol{f}_j$ then we may compute the torsion by

$$(5a) \qquad d(dp) = -du^j d\boldsymbol{f}_j - du^{*j} d\boldsymbol{f}^*{}_j = \theta^{\alpha}{}_j du^j\boldsymbol{f}_{\alpha} + \theta^{\alpha}{}_{j^*} du^{*j}\boldsymbol{f}_{\alpha}$$

in the case of a real displacement or by

$$(5b) \qquad d(dp) = -du^j d\boldsymbol{f}_j = \theta^{\alpha}{}_j du^j\boldsymbol{f}_{\alpha}$$

---

[3] $P^{i}{}_{j} = \partial v^i/\partial u^j$, $P^{i^*}{}_{j^*} = \partial v^{*i}/\partial u^{*j} = P^{*i}{}_{j}$ and $P^{i}{}_{j^*} = 0 = P^{i^*}{}_{j}$ since the v's are analytic functions of the u's. Similar formulas hold for $Q^{\beta}{}_{\alpha}$. Note that $P^{*\beta}{}_{\alpha} = P^{\beta^*}{}_{\alpha^*}$.

in the case of a pure complex displacement. We say that the connection is *without torsion* (in the pure complex sense to be understood) if $d(dp)$ as defined in (5b) is zero.

Having made these definitions we consider the first of the two methods of computing a metric affine connection. In this method we ignore the complex-analytic structure and consider M as a real-analytic manifold with a Riemannian metric (albeit a metric with special form when referred to our complex-analytic local coordinates). Then it is standard that there exists a unique metric affine connection without torsion in the real sense, (5a). In fact we have $\theta^\beta{}_\alpha = B^\beta{}_{\alpha\nu}dz^\nu = g^{\beta\gamma}B_{\alpha\nu,\gamma}dz^\nu$ where

$$B_{\alpha\nu,\gamma} = \tfrac{1}{2}(\partial g_{\alpha\gamma}/\partial u^\nu + \partial g_{\nu\gamma}/\partial u^\alpha - \partial g_{\alpha\nu}/\partial u^\gamma). \tag{6}$$

These formulas simplify somewhat due to the identities (8), section 4, to give:

$$\begin{gathered} B_{ij,k} = 0 = B^*_{i^\bullet j^\bullet,k^\bullet},\ B_{ij^\bullet,k} = \tfrac{1}{2}(\partial g_{j^\bullet k}/\partial u^i - \partial g_{ij^\bullet}/\partial u^k) = B^*_{i^\bullet j,k^\bullet} \\ B_{i^\bullet j,k} = \tfrac{1}{2}(\partial g_{i^\bullet k}/\partial u^i - \partial g_{i^\bullet j}/\partial u^k) = B^*_{ij^\bullet,k^\bullet}, \qquad (7) \\ B_{ij,k^\bullet} = \tfrac{1}{2}(\partial g_{ik^\bullet}/\partial u^j - \partial g_{jk^\bullet}/\partial u^i) = B^*_{i^\bullet j^\bullet,k}. \end{gathered}$$

This affine connection, being the standard one on a real Riemannian manifold defines $d\boldsymbol{v}$ only for real vectors $\boldsymbol{v}$, however it is easily extended to all of W^{2n} since every element of W^{2n} may be uniquely written in the form $\boldsymbol{a} + i\boldsymbol{b}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are real vectors. Thus we define $d(\boldsymbol{a} + i\boldsymbol{b}) = d\boldsymbol{a} + id\boldsymbol{b}$.

Of more interest for complex manifolds is an affine connection which makes essential use of the complex structure, the second method referred to. By analogy with the Riemannian case we would like to think of M as a manifold of n complex dimensions, restrict ourselves to the pure complex vectors of W^{2n}, i. e. to the subspace V^n, and use only the $u^1, \cdots, u^n$ and not $u^{*1}, \cdots, u^{*n}$ to obtain thus the forms $\theta^j{}_i = C^j{}_{ik}du^k$ by which we define $d\boldsymbol{f}_i = \theta^j{}_i\boldsymbol{f}_j$ for the basis vectors $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n$ of V^n. Given such a connection we could extend it to V^{*n}, using $d\boldsymbol{f}^*{}_i = [d\boldsymbol{f}_i]^*$ as definition of $d\boldsymbol{f}^*{}_i$, and hence by linearity to W^{2n}. The condition that the connection be metric gives:

$$\begin{gathered} d(\boldsymbol{f}_i \cdot \boldsymbol{f}^*{}_j) = d\boldsymbol{f}_i \cdot \boldsymbol{f}^*{}_j + \boldsymbol{f}_i \cdot d\boldsymbol{f}^*{}_j \quad \text{or} \\ dg_{ij^\bullet} = \theta^k{}_i g_{kj^\bullet} + \theta^{*k}{}_j g_{ik^\bullet} \quad \text{whence} \\ C_{ij,k^\bullet} = \partial g_{ik^\bullet}/\partial u^j, \qquad C^k{}_{ij} = g^{kl^*}\partial g_{il^\bullet}/\partial u^i \qquad (8) \\ \text{and following } d\boldsymbol{f}^*{}_i = [d\boldsymbol{f}_i]^* \text{ we define} \\ C^*_{ij,k^\bullet} = \partial g_{i^\bullet k}/\partial u^{*i} \text{ and } C^{k^*}{}_{i^\bullet j^\bullet} = C^{*k}{}_{ij}. \text{ All other } C^\gamma{}_{\alpha\beta} \text{ are zero.} \end{gathered}$$

In general the two connections arrived at here are not the same. The

condition that they coincide is exactly that the connection be without torsion (5b). This is equivalent to the Kähler condition on the metric: $\partial g_{ik^*}/\partial u^j = \partial g_{jk^*}/\partial u^i$. However, the main interest below is in a preliminary investigation of this connection when the torsion does not vanish, i. e. the non-Kähler case.

7. The bundle of frames on M. A *frame* $p\boldsymbol{e}_1 \cdots \boldsymbol{e}_n \boldsymbol{e}^*_1 \cdots \boldsymbol{e}^*_n$ at a point p of a Hermitian manifold M will mean an ordered set $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n, \boldsymbol{e}^*_1, \cdots, \boldsymbol{e}^*_n$ of $2n$ vectors of W^{2n}, the vectors $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ in V^n, with order coherent with that orientation of W^{2n} induced by the orientation on M, and satisfying $\boldsymbol{e}_i \cdot \boldsymbol{e}^*_j = \delta_{ij}$. The collection of all such frames at all points of M constitutes the bundle of frames B, a fibre bundle over M with the unitary group $U(n, C)$ as structure group. Since each frame at p is uniquely determined by the vectors $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$, this bundle is, except for notation, the same as that defined by Chern [3]. If $u^1, \cdots, u^n$ are local coordinates in a neighborhood U of M and $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_{2n}$ the corresponding field of natural bases of W^{2n}, $\phi^1, \cdots, \phi^{2n}$ their duals, then the vectors of a frame are given by $\boldsymbol{e}_j = X^k{}_j \boldsymbol{f}_k$ and $\boldsymbol{e}^*{}_j = X^{*k}{}_j \boldsymbol{f}^*{}_k$ or, i. e., $\boldsymbol{e}_\alpha = X^\beta{}_\alpha \boldsymbol{f}_\beta$, with $X^{j^*}{}_i = 0 = X^j{}_{i^*}$. With change of coordinates the $X^\beta{}_\alpha$ transform contravariantly in the upper index. The duals ϵ^j, ϵ^{*j} to $\boldsymbol{e}^j$, $\boldsymbol{e}^{*j}$ are given by $\epsilon^j = Y^j{}_k \phi^k$, $\epsilon^{*j} = Y^{*j}{}_k \phi^{*k}$ or $\epsilon^\alpha = Y^\alpha{}_\beta \phi^\beta$, where $Y^j{}_i X^k{}_j = \delta_{ik} = X^j{}_i Y^k{}_j$, and so we also have $\boldsymbol{f}_j = Y^k{}_j \boldsymbol{e}_k$, $\boldsymbol{f}^*{}_j = Y^{*k}{}_j \boldsymbol{e}^*{}_k$, and so on. The variables $u^1, \cdots, u^n, X^j{}_i, X^{*j}{}_i$ give a system of local coordinates in $\pi^{-1}(U)$ in B, where $\pi : B \to M$ is the natural projection, although the variables $X^j{}_i$ are not independent but satisfy the relations $X^k{}_l X^{*l}{}_j g_{kl^*} = \delta_{ij}$ whence B has complex dimension $n(n+1)/2$ and real dimension $n(n+1)$.

At this point we introduce one of the more important notions in the paper, namely what will be loosely spoken of as *referring* a tensor of M to B. Arguments given in section 5 allow us the convenience of restricting ourselves to pure tensors. To be definite we will consider a tensor $\boldsymbol{T}$ of type $(2, 1)$ having components $T^{ij^*}{}_k$ in the local coordinates $u^1, \cdots, u^n$.

$$\boldsymbol{T} = T^{ij^*}{}_k \boldsymbol{f}_i \otimes \boldsymbol{f}^*{}_j \otimes \phi^k. \tag{1}$$

Relative to the basis of W^{2n} determined by a frame $p\boldsymbol{e}_1 \cdots \boldsymbol{e}_n \boldsymbol{e}^*{}_1 \cdots \boldsymbol{e}^*{}_n$ we have

$$\boldsymbol{T} = \tilde{T}^{lm^*}{}_n \boldsymbol{e}_l \otimes \boldsymbol{e}^*{}_m \otimes \epsilon^n \tag{2}$$

where, of course $\tilde{T}^{lm^*}{}_n$ are given by

$$\tilde{T}^{lm^*}{}_n = T^{ij^*}{}_k Y^l{}_i Y^{*m}{}_j X^k{}_n \tag{3}$$

determining thus n^3 functions defined in $\pi^{-1}(U) \subset B$. It is easy to see that the functions $\tilde{T}^{lm^*}{}_n$ as defined in B are actually independent of the local coordinates on M used on the right side of (3) and thus are functions on all of B, depending in fact only on the tensor $\boldsymbol{T}$ itself. They will be called the *components of* $\boldsymbol{T}$ *in* B. They may be considered as components of $\boldsymbol{T}$ in a sort of universal coordinate system defined over all of M, and a coordinate system for which the vector basis is moreover unit orthogonal throughout. We shall see how it is possible to use these components, functions on B, just as we use the components of $\boldsymbol{T}$ in a local coordinate system; thus "referring" the tensor calculus of M to B. This is a great advantage especially for global considerations as one is freed from the nuisance of local coordinates and, moreover, since the components of $\tilde{T}^{lm^*}{}_n$ are with respect to unit orthogonal bases, it is possible to raise or lower indices merely by placing a bar over the index moved. Thus $\tilde{T}^{lm^*}{}_n$ corresponds to $\tilde{T}^l{}_{nm}$ and to $\tilde{T}^{lm^*n^*}$, etc. This is the method of Chern mentioned in section 1. Below its ramifications are worked out for the case of complex-analytic manifolds and a few applications are given.

In the same fashion that tensors may be referred to B, tensorial forms may also be referred to components in B. Let $\Omega^j{}_i$ be local components of a pure tensorial form of order p and type $(1,1)$; then $\Omega^j{}_i$ has as components in B the n^2 forms $\tilde{\Omega}^j{}_i$ on B determined in local coordinates $u^i, X^j{}_i, X^{*j}{}_i$ by

$$\tilde{\Omega}^j{}_i = \Omega^l{}_k Y^j{}_l X^k{}_i. \tag{4}$$

In particular the complex displacement dp referred to in section 6 is given by $dp = du^i \boldsymbol{f}_i$, the du^i being therefore tensorial forms of type $(1,0)$ and order 1. They have the forms $\omega^i = Y^i{}_j du^j$ as components in B, and, since $dp \cdot dp^* = ds^2$, the differential of arc length, we have $ds^2 = (\omega^1\omega^{*1}) + \cdots + (\omega^n\omega^{*n})$. If $\boldsymbol{v}$ is a vector field on M, then the absolute differential $d\boldsymbol{v}$ is a tensorial form. Let $\boldsymbol{v} = V^\alpha \boldsymbol{f}_\alpha$ in local coordinates then $d\boldsymbol{v}$ has components $DV^\alpha = dV^\alpha + \theta^\alpha{}_\beta V^\beta$ relative to these coordinates. As components in B we get $\widetilde{DV}^\gamma = (dV^\alpha + \theta^\alpha{}_\beta V^\beta) Y^\gamma{}_\alpha$. In particular, if we are given in some neighborhood of M a field of frames $\boldsymbol{e}_i$, then $d\boldsymbol{e}_i = (dX^j{}_i + \theta^j{}_k X^k{}_i)\boldsymbol{f}_j$ which relative to the frames is

$$d\boldsymbol{e}_i = \omega^l{}_i \boldsymbol{e}_l = (dX^j{}_i + \theta^j{}_k X^k{}_i) Y^l{}_j \boldsymbol{e}_l. \tag{5}$$

Now if we allow $X^j{}_i, u^i$ to vary independently then the $\omega^l{}_i$ are forms defined in each coordinate neighborhood of B. But since on every local cross-section of B, i. e. field of frames on M, they define $d\boldsymbol{e}_i$, it is clear that they define

forms on B independent of local coordinates, which assertion is easily checked directly. The $\omega^i, \omega^{*i}, \omega^j{}_i, \omega^{*j}{}_i$ span the space of n forms on B. Relations among them will be considered in section 8. As is already perhaps apparent, the $\omega^j{}_i$ will play the role of $\theta^j{}_i$ in our calculus on B of tensors of M.

Now given as above a tensor $\boldsymbol{T} = T^{ij^*}{}_k \boldsymbol{f}_i \otimes \boldsymbol{f}^*{}_j \otimes \boldsymbol{\phi}^k$ on M, the absolute differential $D\boldsymbol{T}$ of $\boldsymbol{T}$ is a tensorial differential form whose components in the local coordinates are

$$DT^{ij^*}{}_k = dT^{ij^*}{}_k + \theta^i{}_l T^{lj^*}{}_k + \theta^{*j}{}_l T^{il^*}{}_k - \theta^l{}_k T^{ij^*}{}_l. \tag{6}$$

Lifting this tensorial form we get on B

$$\tilde{DT}^{ij}{}_k = DT^{pq^*}{}_r Y^i{}_p Y^{*j}{}_q X^r{}_k. \tag{7}$$

On the other hand the expression by which we will now define $D\tilde{T}^{ij^*}{}_k$, namely

$$D\tilde{T}^{ij^*}{}_k = d\tilde{T}^{ij^*}{}_k + \omega^i{}_l \tilde{T}^{lj^*}{}_k + \omega^j{}_l \tilde{T}^{il^*}{}_k - \tilde{T}^{ij^*}{}_l \omega^l{}_k \tag{8}$$

is easily seen to give the components of $D\boldsymbol{T}$ relative to a field of frames on M and hence, allowing the frames to vary independently as above, to define n^3 forms in B. Further, using $d\tilde{T}^{ij^*}{}_k = d(T^{pq}{}_r Y^i{}_p Y^{*j}{}_q X^r{}_k)$ and (5) we see that

$$D\tilde{T}^{ij^*}{}_k = \tilde{DT}^{ij^*}{}_k. \tag{9}$$

These formulas show that we can refer to B the operation of covariant differentiation merely by using $\omega^j{}_i$ rather than $\theta^j{}_i$. The component differential forms of DT in local coordinates are of the form $T^{ij^*}{}_{k|\alpha} du^\alpha$ where $T^{ij^*}{}_{k|\alpha}$ are the components of the covariant derivatives; and hence $D\tilde{T}^{ij^*}{}_k$ are seen to be given by expressions $D\tilde{T}^{ij^*}{}_k = \tilde{T}^{ij^*}{}_{k|\alpha} \omega^\alpha$ involving only the ω^α's, and not $\omega^j{}_i$'s or their conjugates. Of course $T^{ij^*}{}_{k|l}$ and $T^{ij^*}{}_{k|l^*}$ are tensor components of a tensor, the covariant derivative of $\boldsymbol{T}$, whose components in B are $\tilde{T}^{ij^*}{}_{k|l}$ and $\tilde{T}^{ij^*}{}_{k|l^*}$. We digress to note here that if $\boldsymbol{T}$ is a pure tensor with components $T^{ij^*}{}_k$ relative to a given coordinate system then $\boldsymbol{T}^*$ is also a pure tensor with components $S^{i^*j}{}_{k^*}$ where $S^{i^*j}{}_{k^*} = T^{*ij^*}{}_k$, and from property 3, section 6 of the affine connection it follows that

$$S^{i^*j}{}_{k^*|l} = T^{*ij^*}{}_{k|l^*} \quad \text{and} \quad S^{i^*j}{}_{k^*|l^*} = T^{*ij^*}{}_{k|l}. \tag{10}$$

The content of equation (9) is that for tensors on M the operations of referring to B and of taking covariant derivatives are interchangeable, the forms $\omega^j{}_i$ being used to take covariant derivatives relative to the components of T in B. This leads to an answer to the following question: where are given *a priori* in B the components in B of a tensor of M?

PROPOSITION 4. *Let $\tilde{T}^{ij^*}{}_k$ be n^3 functions on B. These are the components in B of a tensor on M if and only if* [4]

$$d\tilde{T}^{ij^*}{}_k + T^{ij^*}{}_k\omega^i{}_l + \tilde{T}^{il^*}{}_k\omega^{*j}{}_l - \tilde{T}^{ij^*}{}_l\omega^l{}_k \equiv 0(\omega^\alpha). \tag{11}$$

Proof. The necessity is proved above and the sufficiency follows from the sufficiency proof of the following proposition, as does in fact the entire proposition if we regard tensors as tensorial forms of order 0.

PROPOSITION 5. *Necessary and sufficient conditions that a set of n^3 differential forms $\tilde{\Omega}^{ij^*}{}_k$ of order p on B are the components in B of a tensorial form of M are* [4]

$$\tilde{\Omega}^{ij^*}{}_k \in R[\omega^\alpha] \tag{12}$$

and

$$d\tilde{\Omega}^{ij^*}{}_k + \omega^i{}_l\tilde{\Omega}^{lj^*}{}_k + \omega^{*j}{}_l\tilde{\Omega}^{il^*}{}_k - \omega^l{}_k\tilde{\Omega}^{ij^*}{}_l \in R[\omega_\alpha]. \tag{13}$$

Proof. The necessity is a matter of direct verification and is similar to the remarks above for a tensor $\tilde{T}^{ij^*}{}_k$ and so will not be repeated. Before proving the sufficiency we make some preliminary statements about forms of M and B and prove a lemma. We continue to let $\pi: B \to M$ denote the natural map carrying each frame $p\boldsymbol{e}_1 \cdots \boldsymbol{e}_n\boldsymbol{e}^*{}_1 \cdots \boldsymbol{e}^*{}_n$ into its base point p, and let π again denote the induced map of forms on M to forms on B. By considering local coordinates $u^1, \cdots, u^n$ in a neighborhood U on M and $u^1, \cdots, u^n, X^j{}_i, X^{*j}{}_i$ in $\pi^{-1}(U)$ on B we see that the image in B of a form of M must involve only du^α and not $dX^j{}_i$ or $dX^{*j}{}_i$, or, using $\omega^\alpha, \omega^j{}_i, \omega^{*j}{}_i$, then only ω^α and not $\omega^j{}_i$ or $\omega^{*j}{}_i$. Thus a necessary condition that ϕ, a form of B, be the image, $\phi = \pi(\psi)$, of a form of M is $\phi \in R[\omega^\alpha]$. Further, in the local coordinates of $\pi^{-1}(U)$ any form $\pi(\psi)$ must have as coefficients, when expressed relative to du^α, functions of $u^1, \cdots, u^n, u^{*1}, \cdots, u^{*n}$ for which an equivalent condition is that the exterior derivative involves only the du^α's and not $dX^j{}_i$, $dX^{*j}{}_i$ so that the following lemma holds.

LEMMA. *A form ϕ of B is the image under π of a form ψ of M if and only if*

$$\phi \in R[\omega^\alpha], \qquad d\phi \in R[\omega^\alpha]. \tag{14}$$

Now consider again $\tilde{\Omega}^{ij^*}{}_k$ satisfying (13). A tensorial form on M which has

[4] A differential form ϕ of B will be said to be congruent to zero mod ω^α, $\phi \equiv 0(\omega^\alpha)$, if ϕ is in the ideal generated by the ω^α's, $\alpha = 1, \cdots, 2n$, whereas $\phi \in R[\omega^\alpha]$ will denote that ϕ is a polynomial in the ω^α's. For ϕ linear the two are equivalent. As previously remarked restriction to tensors of type (2, 1) is for definiteness only.

components $\tilde{\Omega}^{ij^*}{}_k$ in B would in local coordinates in $\pi^{-1}(U)$ on B necessarily be given by

$$\Omega^{mn^*}{}_l = X^m{}_j X^{*n}{}_k Y^i{}_l \tilde{\Omega}^{jk^*}{}_i \tag{15}$$

and if the coefficients of the forms $\Omega^{mn^*}{}_l$ are functions of u^i, u^{*i} alone, then it is easy to see that $\Omega^{mn^*}{}_l$ are in fact the components in the coordinates u^i of a tensorial form on M. By the lemma this holds if $d\Omega^{mn^*}{}_l \equiv R[\omega^a]$. Now

$$\begin{aligned} d\Omega^{mn^*}{}_l = {} & X^m{}_j X^{*n}{}_k Y^i{}_l d\tilde{\Omega}^{jk^*}{}_i + dX^m{}_j X^{*n}{}_k Y^i{}_l \tilde{\Omega}^{jk^*}{}_i \\ & + X^m{}_j dX^{*n}{}_k Y^i{}_l \tilde{\Omega}^{jk^*}{}_i + X^m{}_j X^{*n}{}_k dY^i{}_l \tilde{\Omega}^{jk^*}{}_i. \end{aligned}$$

Using (13) and the equations (5)

$$dX^j{}_i = \omega^k{}_i X^j{}_k - \theta^j{}_k X^k{}_i \qquad dY^j{}_i = -Y^k{}_i \omega^j{}_k + Y^j{}_k \theta^k{}_i$$

we get

$$d\Omega^{mn^*}{}_l = -\theta^m{}_k \Omega^{kn^*}{}_l - \theta^{*n}{}_k \Omega^{mn^*}{}_l + \theta^k{}_l \Omega^{mn^*}{}_k \equiv R[\omega^a]$$

which completes the proof.

8. Equations of structure. Chern [3] has proved that given a Hermitian manifold M, its bundle of frames B and the forms $\omega^1, \cdots, \omega^n$ defined locally by $\omega^i = Y^i{}_j du^j$ in the notation of the previous section, then there exists a unique set of forms $\omega^i{}_j$ on B satisfying

$$\omega^j{}_i + \omega^{*i}{}_j = 0 \tag{1}$$

$$d\omega^i - \omega^k \omega^i{}_k = A^i{}_{jk} \omega^j \omega^k, \qquad A^i{}_{jk} + A^i{}_{kj} = 0. \tag{2}$$

If we denote by Ω^i the forms $A^i{}_{jk} \omega^j \omega^k$ he has further shown that the forms $\Omega^j{}_i$ defined by

$$\Omega^j{}_i = d\omega^j{}_i - \omega^k{}_i \omega^j{}_k \tag{3}$$

are of the form [5]

$$\Omega^j{}_i = R_i{}^j{}_{kl^\bullet} \omega^k \omega^{*l}, \qquad R_{ij^\bullet kl^\bullet} = R^*{}_{ji^\bullet lk^\bullet}. \tag{4}$$

These he has called the *curvature* forms. By taking the derivatives of (2) and (3) and using these equations themselves one gets

$$d\Omega^i + \omega^i{}_j \Omega^j = \omega^j \Omega^i{}_j \tag{5}$$

$$d\Omega^j{}_i - \omega^j{}_k \Omega^k{}_i - \omega^k{}_i \Omega^j{}_k = 0 \tag{6}$$

whence, since the right hand side of each equation is in $R[\omega^a]$, we have

[5] We have altered slightly the notation of Chern writing as $R_{ij^\bullet kl^\bullet}$ the quantity for which he has written $-R_{ij^\bullet l^\bullet k^\bullet}$.

from Proposition 5 that Ω^i, $\Omega^j{}_i$ are the components in B of tensorial forms of M. The corresponding tensorial forms of M of course are the torsion and curvature forms as will be shown.

First we must show that the $\omega^j{}_i$ whose existence, uniqueness, and role in the local equivalence problem are demonstrated in [3] are the same as those we defined in section 7 by

$$\omega^j{}_i = (dX^l{}_i + \theta^l{}_k X^k{}_i) Y^j{}_l \tag{7}$$

that is, we must show that our $\omega^j{}_i$ satisfy (1) and (2). Then it will appear, by solving for $\theta^l{}_k$ in the equation above, that the $\theta^l{}_k$ we determined by assuming the connection to be metric, to satisfy $d\boldsymbol{f}^*{}_j = [d\boldsymbol{f}_j]^*$, and to be of the form $\theta^j{}_i = C^j{}_{ik} du^k$ are the same essentially as those determined by (1) and (2) above, which throws more light on why this second type of affine connection is natural for Hermitian manifolds. We have

$$d\omega^i = dY^i{}_j du^j = -Y^i{}_j dX^j{}_k \omega^k \tag{8}$$

and using (7) we get

$$d\omega^i = \omega^k \omega^i{}_k - Y^i{}_j X^l{}_k g^{jp^*} \partial g_{lp^*}/\partial u^q X^q{}_i \omega^h \omega^k, \quad \text{hence}$$

$$d\omega^i = \omega^k \omega^i{}_k + A^i{}_{hk} \omega^h \omega^k \quad \text{with} \tag{9}$$

$$A^i{}_{hk} = \tfrac{1}{2}(Y^i{}_j X^l{}_h X^q{}_k - Y^i{}_j X^l{}_k X^q{}_h) g^{jp^*} \partial g_{lr^*}/\partial u^q. \tag{10}$$

Reference to (5b) of section 6 will show the forms $\Omega^i = A^i{}_{hk} \omega^h \omega^k$ to be exactly the components in B of the torsion $\theta^i{}_j du^j = C^i{}_{jk} du^k du^j$ as defined there. It remains to show that the forms $\omega^j{}_i$ defined by (7) satisfy (1). We have

$$\omega^j{}_i + \omega^{*i}{}_j = (dX^k{}_i + X^l{}_i \theta^k{}_l) Y^j{}_k + (dX^{*k}{}_j + X^{*l}{}_j \theta^{*k}{}_l) Y^{*i}{}_k. \tag{11}$$

The condition $\boldsymbol{e}_i \cdot \boldsymbol{e}^*{}_j = \delta_{ij}$ may be written

$$X^k{}_i X^{*l}{}_j g_{kl^*} = \delta_{ij}. \tag{12}$$

Differentiating we get

$$dX^k{}_i X^{*l}{}_j g_{kl^*} + X^k{}_i dX^{*l}{}_j g_{kl^*} + X^k{}_i X^{*l}{}_j dg_{kl^*} = 0. \tag{13}$$

If we use $g_{kl^*} = Y^p{}_k Y^{*p}{}_l$, this becomes

$$dX^k{}_i Y^j{}_k + dX^{*l}{}_j Y^{*i}{}_l + X^k{}_i X^{*l}{}_j dg_{kl^*} = 0, \tag{14}$$

and we may write

$$\omega^j{}_i + \omega^{*i}{}_j = -X^m{}_i X^{*n}{}_j dg_{mn^*} + X^l{}_i Y^j{}_k \theta^k{}_l + X^{*l}{}_j Y^{*i}{}_k \theta^{*k}{}_l. \tag{15}$$

Then substituting for $\theta^k{}_l$, $\theta^{*k}{}_l$ the expressions found in section 6, (8), we get, using $g^{jk^*} = X^j{}_p X^{*k}{}_p$,

$$\omega^j{}_i + \omega^{*i}{}_j = 0 \tag{1}$$

as desired.

If we denote by $\Lambda^k{}_i$ the curvature forms in the local coordinates u^j, u^{*j} we can show in much the same fashion as above that if these are given locally in M by the formulas

$$\Lambda^k{}_i = d\theta^k{}_i - \theta^h{}_i\theta^k{}_h \tag{16}$$

then the components in B of these forms are in fact $\Omega^k{}_i$ as defined in (3). The equations (16) allow us to compute $\Lambda^k{}_i$ explicitly. We have from section 6, (8),

$$\theta^k{}_i = C^k{}_{il}du^l, \qquad C^k{}_{il} = g^{kj^*}\,\partial g_{ij^*}/\partial u^l \tag{17}$$

so that

$$\Lambda^k{}_i = d(C^k{}_{il}du^l) - C^h{}_{im}C^k{}_{hl}du^m du^l$$

$$\Lambda^k{}_i = (\partial C^k{}_{il}/\partial u^m - C^h{}_{im}C^k{}_{hl})\,du^m du^l + \partial C^k{}_{il}/\partial u^{*m}\,du^{*m}du^l. \tag{18}$$

Using the notation

$$\Lambda^k{}_i = R_i{}^k{}_{ml}du^m du^l + R_i{}^k{}_{lm^*}\,du^l du^{*m} \tag{19}$$

where $R_i{}^k{}_{ml} + R_i{}^k{}_{lm} = 0$ we have

$$R_i{}^k{}_{ml} = \tfrac{1}{2}(\partial C^k{}_{il}/\partial u^m - \partial C^k{}_{im}/\partial u^l - C^h{}_{im}C^k{}_{hl} + C^h{}_{il}C^k{}_{hm}),$$

$$R_i{}^k{}_{lm^*} = -\partial C^k{}_{il}/\partial u^{*m}. \tag{20}$$

Substituting from (17) in (20) and using $d(g^{ij^*}g_{kj^*}) = 0$ we get

$$R_i{}^k{}_{ml} = 0 \quad \text{and} \tag{21}$$

$$R_i{}^k{}_{ml^*} = -\partial^2 g_{ij^*}/\partial u^l \partial u^{*m}\, g^{kj^*} - (\partial g_{ij^*}/\partial u^l)\,(\partial g^{kj^*}/\partial u^{*m}) \tag{22}$$

which may be written

$$R_{ik^*ml} = 0 \tag{23}$$

$$R_{ik^*ml^*} = -\partial^2 g_{ik^*}/\partial u^m \partial u^{*l} + g^{pq^*}\,(\partial g_{pk^*}/\partial u^m)\,(\partial g_{iq^*}/\partial u^{*l}). \tag{24}$$

10. The Bianchi identities. We have defined Ω^i, the torsion forms, by

$$\Omega^i = A^i{}_{jk}\omega^j\omega^k \tag{1}$$

and have noted section 8, (5) that

$$d\Omega^i - \omega^i{}_j\Omega_j = \omega_j\Omega^i{}_j. \tag{2}$$

If we take the exterior derivative of both sides of equation (1) and substitute for $d\Omega^i$ in (2) we are led to

$$A^i{}_{jk|l}\omega^l\omega^j\omega^k + A^i{}_{jk|l^*}\omega^{*l}\omega^j\omega^k + 2A^i{}_{hk}A^h{}_{lj}\omega^l\omega^j\omega^k = R_{ji^*kl^*}\omega^{*l}\omega^j\omega^k \tag{3}$$

which gives us the *Bianchi identities*

$$R_{ji^*kl^*} - R_{ki^*jl^*} = 2A^i{}_{jk|l^*} \tag{4a}$$

$$A^i{}_{jk|l} - A^i{}_{lk|j} - A^i{}_{jl|k} + 2(A^i{}_{hk}A^h{}_{lj} - A^i{}_{hl}A^h{}_{kj} - A^i{}_{hj}A^h{}_{lk}) = 0. \tag{4b}$$

Concerning (4a) we remark that if $A^i{}_{jk|l^*} = 0$ then the curvature tensor has the same symmetry properties as in the case of a Kähler metric. This seems to be one of two rather natural conditions that can be imposed on the torsion, the other being $A^h{}_{jh} = 0$. Unfortunately examples of non-Kähler manifolds are not plentiful, and at present there seems to be no example of a compact manifold for which either of these conditions is satisfied. In particular if $A^i{}_{jk|l^*} = 0$ we may conclude that the three "Ricci tensors" coincide:

$$\begin{aligned} R_{ij^*} &= R_{hh^*ij^*} \\ S_{ij^*} &= R_{ih^*hj^*} \\ T_{ij^*} &= R_{ij^*hh^*}. \end{aligned} \tag{5}$$

On the other hand, if $A^h{}_{jh} = 0$ then $R_{ij^*} = S_{ij^*}$. That these three tensors need not be the same is shown by the following example.

We consider the manifold $M = S^1 \times S^3$, the Cartesian product of a 1-sphere and a 3-sphere. Hopf [7] has shown that this space is a complex-analytic manifold as follows. Let u^1, u^2 be complex coordinates in the 4-dimensional Euclidean space, and G the discontinuous group on $E^4 - \{(0,0)\}$ generated by the transformation $u'^1 = 2u^1$; $u'^2 = 2u^2$. Identification of points congruent under G gives $S^1 \times S^3$ plus a complex-analytic structure inherited from E^4. A natural metric is given by

$$ds^2 = (du^1 du^{*1} + du^2 du^{*2})/\lambda, \qquad \lambda = u^1 u^{*1} + u^2 u^{*2} \tag{6}$$

so that

$$g_{ij^*} = \delta_{ij}/\lambda \quad \text{and} \quad g^{ij^*} = \lambda \delta^{ij}. \tag{7}$$

An easy computation then gives

$$\begin{aligned} C^i{}_{jk} &= -\delta_{ij}u^{*k}/\lambda, & A^i{}_{jk} &= (\delta_{ik}u^{*j} - \delta_{ij}u^{*k})/2\lambda \\ R_{ij^*kl^*} &= (\delta_{ij}\delta_{kl}\lambda - \delta_{ij}u^{*k}u^l)/\lambda^3, & R_{ij^*} &= 2(\delta_{ij}\lambda - u^{*i}u^j)/\lambda^3 \\ S_{ij^*} &= (\delta_{ij}\lambda - u^{*i}u^j)/\lambda^3, & T_{ij^*} &= \delta_{ij}/\lambda^2. \end{aligned} \tag{8}$$

11. Closed forms of M. We will call any form on B which is the image under π of a form on M *a form of* M. Necessarily any such form ϕ satisfies

$$\phi \in R[\omega^a]. \tag{1}$$

Let ϕ be a form satisfying (1). From section 5, ϕ is a sum of pure forms, to which we shall then restrict ourselves:

$$\phi = \lambda_{i_1 \dots i_p j^*_1 \dots j^*_q} \omega^{i_1} \cdot \cdot \cdot \omega^{i_p} \omega^{*j_1} \cdot \cdot \cdot \omega^{*j_q},$$

where $\lambda_{i_1 \dots i_p j^*_1 \dots j^*_q}$ is skew in $i_1 \cdot \cdot \cdot i_p$ and skew in $j_1 \cdot \cdot \cdot j_q$. We evaluate $d\phi$:

$$\begin{aligned}(2) \qquad d\phi = \{ & d\lambda_{i_1 \dots i_p j^*_1 \dots j^*_q} - \sum_{r=1}^{p} \lambda_{i_1 \dots k \dots i_p j^*_1 \dots j^*_q} \omega^k{}_{i_r} \\ & - \sum_{s=1}^{q} \lambda_{i_1 \dots i_p j^*_1 \dots k^* \dots j^*_q} \omega^{*k}{}_{j_s} - \sum_{r=1}^{p} \lambda_{i_1 \dots k \dots i_p j^*_1 \dots j^*_q} A^k{}_{i_r l} \omega^l \\ & - \sum_{s=1}^{q} \lambda_{i_1 \dots i_p j^*_1 \dots k^* \dots j^*_q} A^{*k}{}_{j_s l} \omega^{*l} \} \omega^{i_1} \cdot \cdot \cdot \omega^{i_p} \omega^{*j_1} \cdot \cdot \cdot \omega^{*j_q}.\end{aligned}$$

If $d\phi = 0$ then the term in braces must be a linear combination of ω^α's i. e. $\equiv 0(\omega^\alpha)$ and hence in particular the terms in braces not involving $A^k{}_{ij}, A^{*k}{}_{ij}$ are $\equiv 0(\omega^\alpha)$, i. e., $\lambda_{i_1 \dots i_p j^*_1 \dots j^*_q}$ is a pure tensor of M:

$$\begin{aligned}(3) \qquad & d\lambda_{i_1 \dots i_p j^*_1 \dots j^*_q} - \sum_{r=1}^{p} \lambda_{i_1 \dots k \dots i_p j^*_1 \dots j^*_q} \omega^k{}_{i_r} - \sum_{s=1}^{q} \lambda_{i_1 \dots i_p j^*_1 \dots k^* \dots j^*_q} \omega^{*k}{}_{j_s} \\ & = \lambda_{i_1 \dots i_p j^*_1 \dots j^*_q | l} \omega^l + \lambda_{i_1 \dots i_p j^*_1 \dots j^*_q | l^*} \omega^{*l}.\end{aligned}$$

For any function λ on B (or on M) $d\lambda$ splits into the unique sum $d_1\lambda + d_2\lambda$ where $d_1\lambda$ is a form in ω^i, $\omega^j{}_i$ and $d_2\lambda$ is a form in ω^{*i}, $\omega^{*j}{}_i$ and (3) may be written

$$\begin{aligned}(3a) \qquad & d_1\lambda_{i_1 \dots i_p j^*_1 \dots j^*_q} - \sum_{r=1}^{p} \lambda_{i_1 \dots k \dots i_p j^*_1 \dots j^*_q} \omega^k{}_{i_r} \\ & = \lambda_{i_1 \dots i_p j^*_1 \dots j^*_q | l} \omega^l \\ (3b) \qquad & d_2\lambda_{i_1 \dots i_p j^*_1 \dots j^*_q} - \sum_{s=1}^{q} \lambda_{i_1 \dots i_p j^*_1 \dots k^* \dots j^*_q} \omega^{*k}{}_{j_s} \\ & = \lambda_{i_1 \dots i_p j^*_1 \dots j^*_q | l^*} \omega^{*l}.\end{aligned}$$

If $d\phi = 0$, we see from Lemma 1, section 7, that $\phi = \pi\psi$, i. e. ϕ is a form of M. Moreover since d and π are interchangeable and π is an isomorphism into it follows that $d\psi = 0$. If we substitute (3a) and 3b) in (2) we get the conditions

$$\begin{aligned}(4a) \qquad & \lambda_{i_1 \dots i_p j^*_1 \dots j^*_q | l} - \sum_{r=1}^{p} \lambda_{i_1 \dots l \dots i_p j^*_1 \dots j^*_q | i_r} \\ & - 2 \sum_{r=1}^{p} \lambda_{i_1 \dots k \dots i_p j^*_1 \dots j^*_q} A^k{}_{i_r l} + \sum_{t \neq r}^{p} \sum \lambda_{i_1 \dots l \dots k \dots i_p j^*_1 \dots j^*_q} A^k{}_{i_r i_t} = 0 \\ (4b) \qquad & \lambda_{i_1 \dots i_p j^*_1 \dots j^*_q | l^*} - \sum_{s=1}^{p} \lambda_{i_1 \dots i_p j^*_1 \dots l^* \dots j^*_q | j^*_s} \\ & - 2 \sum_{s=1}^{q} \lambda_{i_1 \dots i_p j^*_1 \dots k^* \dots j^*_q} A^{*k}{}_{j_s l} + \sum_{t \neq s}^{q} \sum \lambda_{i_1 \dots i_p j^*_1 \dots l^* \dots k^* \dots j^*_q} A^{*k}{}_{j_s j_t} = 0.\end{aligned}$$

Giving in summary

PROPOSITION 6. *If ϕ, a pure r-form of B of type (p, q) is in $R[\omega^a]$, then $d\phi = 0$ implies ϕ is a form of M,* i. e. *$\phi = \pi\psi$, and that $d\psi = 0$. The necessary and sufficient conditions that a form of M be closed are* (4a) *and* (4b).

12. The forms $*\theta$ and $\mathfrak{G}\theta$. We let

$$\theta = \lambda\omega^{i_1} \cdots \omega^{i_p}\omega^{*j_1} \cdots \omega^{*j_q} \tag{1}$$

for a fixed set of distinct indices $i_1 \cdots i_p$ and a fixed set of distinct indices $j_1 \cdots j_q$, and define

$$*\theta = (-1)^{(n-p)q} i^n \lambda^* \omega^{k_1} \cdots \omega^{k_{n-p}}\omega^{*l_1} \cdots \omega^{*l_{n-q}} \tag{2}$$

where $i_1, \cdots, i_p, k_1, \cdots, k_{n-p}$ and $j_1 \cdots j_q, l_1, \cdots, l_{n-q}$ form even permutations of $1, \cdots, n$. Then

$$\theta*\theta = \lambda\lambda^* i^n \omega^1 \cdots \omega^n \omega^{*1} \cdots \omega^{*n} = \lambda\lambda^*\Theta \tag{3}$$

where we denote the volume element by Θ

$$\Theta = i^n \omega^1 \cdots \omega^n \omega^{*1} \cdots \omega^{*n}. \tag{4}$$

We define $*\theta$ for general forms by linearity. It is worth noting that

$$**\theta = (-1)^r\theta, \qquad r = p + q. \tag{5}$$

Given a pure form θ of type (p, q) we can write for $*\theta$

$$\begin{aligned} &*(\lambda_{i_1 \cdots i_p j^*_1 \cdots j^*_q} \omega^{i_1} \cdots \omega^{i_p}\omega^{*j_1} \cdots \omega^{*j_q}) \\ &= i^n(-1)^{(n-p)q} \frac{1}{(n-p)!} \frac{1}{(n-q)!} \epsilon_{i_1 \cdots i_p k_1 \cdots k_{n-p}} \epsilon^*_{j_1 \cdots j_q l_1 \cdots l_{n-q}} \\ &\qquad \lambda^*_{i_1 \cdots i_p j^*_1 \cdots j^*_q} \omega^{k_1} \cdots \omega^{k_{n-p}}\omega^{*l_1} \cdots \omega^{*l_{n-q}} \end{aligned} \tag{6}$$

where $\epsilon_{p_1 \cdots p_n}$ is zero unless $p_1 \cdots p_n$ form a permutation of $1, \cdots, n$ and then is $+1$ if it is an even permutation, -1 if it is an odd permutation.

If we introduce the notation

$$\begin{aligned} &\mu_{k_1 \cdots k_{n-p} l^*_1 \cdots l^*_{n-q}} \\ &= \frac{1}{(n-p)!} \frac{1}{(n-q)!} \epsilon_{i_1 \cdots i_p k_1 \cdots k_{n-p}} \epsilon^*_{j_1 \cdots j_q l_1 \cdots l_{n-q}} \lambda^*_{i_1 \cdots i_p j^*_1 \cdots j^*_q}, \end{aligned} \tag{7}$$

we have

$$\begin{aligned} &\lambda_{i_1 \cdots i_p j^*_1 \cdots j^*_q} \\ &= \frac{1}{p!} \frac{1}{q!} \epsilon_{i_1 \cdots i_p k_1 \cdots k_{n-p}} \epsilon^*_{j_1 \cdots j_q l_1 \cdots l_{n-q}} \mu^*_{k_1 \cdots k_{n-p} l^*_1 \cdots l^*_{n-q}}. \end{aligned} \tag{8}$$

This notation allows us to write

$$(9) \qquad *\theta = i^n(-1)^{(n-p)q}\mu_{k_1\dots k_{n-p}l^*_1\dots l^*_{n-q}}\omega^{k_1}\cdots\omega^{k_{n-p}}\omega^{*l_1}\cdots\omega^{*l_{n-q}},$$

and thus using our computation for d, and assuming $\lambda_{i_1\dots i_p j^*_1\dots j^*_q}$ are components of a tensor of M, i. e. θ a form of M, we get

$$(10) \qquad d*\theta = i^n(-1)^{(n-p)q}\{\mu_{k_1\dots k_{n-p}l^*_1\dots l^*_{n-q}|h}\omega^h + \mu_{k_1\dots k_{n-p}l^*_1\dots l^*_{n-q}|h^*}\omega^{*h}$$
$$+\sum_{r=1}^{n-p}\mu_{k_1\dots t\dots k_{n-p}l^*_1\dots l^*_{n-q}}A^t{}_{hk_r}\omega^h$$
$$+\sum_{s=1}^{n-q}\mu_{k_1\dots k_{n-p}l^*_1\dots t^*\dots l^*_{n-q}}A^{*t}{}_{hl_s}\omega^{*h}\}\omega^{k_1}\cdots\omega^{k_{n-p}}\omega^{*l_1}\cdots\omega^{*l_{n-q}}.$$

Following this by a straightforward but long computation we compute $\delta\theta = *d*\theta$ to be

$$(11) \qquad \delta\theta = p(\lambda_{hi_1\dots i_{p-1}j^*_1\dots j^*_q|h^*} + 2\lambda_{hi_1\dots i_{p-1}j^*_1\dots j^*_q}A^{*k}{}_{hk}$$
$$-\sum_{r=1}^{p-1}\lambda_{hi_1\dots k\dots i_{p-1}j^*_1\dots j^*_q}A^{*i_r}{}_{hk})\omega^{i_1}\cdots\omega^{i_{p-1}}\omega^{*j_1}\cdots\omega^{*j_q}$$
$$+(-1)^p q(\lambda_{i_1\dots i_p h^*j^*_1\dots j^*_{q-1}|h} + 2\lambda_{i_1\dots i_p h^*j^*_1\dots j_{q-1}}A^k{}_{hk}$$
$$-\sum_{s=1}^{q-1}\lambda_{i_1\dots i_p h^*j^*_1\dots k^*\dots j^*_{q-1}}A^{j_s}{}_{hk})\omega^{i_1}\cdots\omega^{i_p}\omega^{*j_1}\cdots\omega^{*j_{q-1}}.$$

The condition that $\delta\theta = 0$ is then that each term in parenthesis is zero. We remark that since it is clear that * commutes with π, i. e. $*\pi\psi = \pi^*\psi$, and since we have noted that $d\pi = \pi d$, we have

PROPOSITION 7. *If $\theta = \pi\psi$, then $\delta\theta = 0$ if and only if $\delta\psi = 0$. Moreover if $\theta \in R[\omega^\alpha]$, then $\delta\theta = 0$ is sufficient to ensure that θ is a form of M.*

13. The interchange formulas. We consider the components in B of a pure tensor of M, $\lambda_{i_1\dots i_p j^*_1\dots j^*_q}$, these components being functions in B. The condition that these are components of a tensor in M is, according to section 7, Proposition 4,

$$(1) \qquad d\lambda_{i_1\dots i_p j^*_1\dots j^*_q} - \sum_{r=1}^{p}\lambda_{i_1\dots k\dots i_p j^*_1\dots j^*_q}\omega^k{}_{i_r} - \sum_{s=1}^{q}\lambda_{i_1\dots i_p j^*_1\dots k^*\dots j^*_q}\omega^{*k}{}_{j_s}$$
$$= \lambda_{i_1\dots i_p j^*_1\dots j^*_q|\alpha}\omega^\alpha$$

which expresses the fact that the left side is $\equiv 0(\omega^\alpha)$. $\lambda_{i_1\dots i_p j^*_1\dots j^*_q|\alpha}$ are the components in B of the covariant derivative of the tensor. Taking the exterior derivative on both sides of (1) and collecting terms we get

$$
\begin{aligned}
&-\sum_{r=1}^{p}\left(d\lambda_{i_1\ldots k\ldots i_p j^*_1\ldots j^*_q}-\lambda_{i_1\ldots l\ldots i_p j^*_1\ldots j^*_q}\omega^l{}_k\right)\omega^k{}_{i_r}-\sum_{s=1}^{q}\lambda_{i_1\ldots i_p j^*_1\ldots k^*\ldots j^*_q}\Omega^{*k}{}_{j_s}\\
&-\sum_{s=1}^{q}\left(d\lambda_{i_1\ldots i_p j^*_1\ldots k^*\ldots j^*_q}-\lambda_{i_1\ldots i_p j^*_1\ldots l^*\ldots j^*_q}\omega^{*l}{}_k\right)\omega^{*k}{}_{j_s}-\sum_{r=1}^{p}\lambda_{i_1\ldots k\ldots i_p j^*_1\ldots j^*_q}\Omega^k{}_{i_r}\\
&=d\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|\alpha}\omega^\alpha+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k^*}\omega^{*l}\omega^{*k}{}_l\\
&\quad+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k}\omega^l\omega^k{}_l+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k}\Omega^k+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k^*}\Omega^{*k}.
\end{aligned}
\tag{2}
$$

By adding terms whose sum is zero we see that (2) may be written

$$
\begin{aligned}
&\{-d\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|l}-\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k}\omega^k{}_l-\sum_{r=1}^{p}\lambda_{i_1\ldots k\ldots i_p j^*_1\ldots j^*_q|l}\omega^k{}_{i_r}\\
&\quad-\sum_{s=1}^{q}\lambda_{i_1\ldots i_p j^*_1\ldots k^*\ldots j^*_q|l}\omega^{*k}{}_{j_s}\}\omega^l\\
&\{-d\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|l^*}-\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k^*}\omega^{*k}{}_l-\sum_{r=1}^{p}\lambda_{i_1\ldots k\ldots i_p j^*_1\ldots j^*_q|l^*}\omega^k{}_{i_r}\\
&\quad-\sum_{s=1}^{q}\lambda_{i_1\ldots i_p j^*_1\ldots k^*\ldots j^*_q|l^*}\omega^{*k}{}_{j_s}\}\omega^{*l}\\
&\quad=\sum_{r=1}^{p}\lambda_{i_1\ldots k\ldots i_p j^*_1\ldots j^*_q}\Omega^k{}_{i_r}+\sum_{s=1}^{q}\lambda_{i_1\ldots i_p j^*_1\ldots k^*\ldots j^*_q}\Omega^{*k}{}_{j_s}\\
&+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k}\Omega^k+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k^*}\Omega^{*k}.
\end{aligned}
\tag{3}
$$

This gives at once

$$
\begin{aligned}
&\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|lk}\omega^l\omega^k+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|lk^*}\omega^l\omega^{*k}\\
&+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|l^*k}\omega^{*l}\omega^k+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|l^*k^*}\omega^{*l}\omega^{*k}\\
&=\sum_{r=1}^{p}\lambda_{i_1\ldots k\ldots i_p j^*_1\ldots j^*_q}\Omega^k{}_{i_r}+\sum_{s=1}^{q}\lambda_{i_1\ldots i_p j^*_1\ldots k^*\ldots j^*_q}\Omega^{*k}{}_{j_s}\\
&+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k}\Omega^k+\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k^*}\Omega^{*k}.
\end{aligned}
\tag{4}
$$

From this, by equating coefficients, we obtain the formulas for the interchange of the order of covariant differentiation indices:

$$
\begin{aligned}
\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|kl^*}-\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|l^*k}&=\sum_{r=1}^{p}\lambda_{i_1\ldots h\ldots i_p j^*_1\ldots j^*_q}R_{i_r h^* kl^*}\\
&\quad-\sum_{s=1}^{q}\lambda_{i_1\ldots i_p j^*_1\ldots h^*\ldots j^*_q}R_{hj^*_s kl^*}
\end{aligned}
\tag{5a}
$$

$$
\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|lk}-\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|kl}=2\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|h}A^h{}_{lk}
\tag{5b}
$$

$$
\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|l^*k^*}-\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|k^*l^*}=2\lambda_{i_1\ldots i_p j^*_1\ldots j^*_q|h^*}A^{*h}{}_{lk}.
\tag{5c}
$$

Note that (5a) is the same as in the Kähler case, i.e., the case where $\Omega^i=0$. Of course, since covariant differentiation is a linear operation, and since the order of the indices was not used, these formulas may be extended

to any tensor, including those with contravariant indices, since it is easy to show that the lowering and raising of indices commutes with differentiation.

14. Bochner's lemma.[6] We now prove a slightly extended version of the lemma used by Bochner [2]. In the sequel the lemma will be applied in the same fashion as in his paper only to the more general case of a Hermitian metric with torsion. The presence of torsion, it will be seen, enormously complicates the formulas, so that only the simplest cases will be discussed.

Let Φ be a real-valued scalar function on M, defining then a function, also denoted Φ, on B. We have

$$d\Phi = \Phi_k\omega^k + \Phi_{k\bullet}\omega^{*k}, \qquad \Phi_{k\bullet} = \Phi^*{}_k. \tag{1}$$

Our formula for $\delta\theta$ then gives

$$*d*d\Phi = \Phi_{kk\bullet} + 2\Phi_k A^{*h}{}_{kh} + \Phi_{k\bullet k} + 2\Phi_{k\bullet}A^h{}_{kh} \tag{2}$$

where $\Phi_{ij\bullet}$ and $\Phi_{i\bullet j}$ are the second covariant derivatives of Φ and satisfy $\Phi_{ij\bullet} = \Phi_{j\bullet i}$ and $\Phi_{ij\bullet} = \Phi^*{}_{i\bullet j}$, whence

$$*d*d\Phi = 2(\Phi_{kk\bullet} + \Phi_k A^{*h}{}_{kh} + \Phi_{k\bullet}A^h{}_{kh}). \tag{3}$$

If we take $*$ of (3) we get

$$d*d\Phi = -2i^n(\Phi_{kk\bullet} + \Phi_k A^{*h}{}_{kh} + \Phi_{k\bullet}A^h{}_{kh})\Theta \tag{4}$$

where $\Theta = i^n\omega^1\cdots\omega^n\omega^{*1}\cdots\omega^{*n}$ is the volume element of M. Then by Stokes' Theorem, since the integral of an exact differential over a cycle vanishes,

$$\int_M (\Phi_{kk\bullet} + \Phi_k A^{*h}{}_{kh} + \Phi_{k\bullet}A^h{}_{kh})\Theta = 0. \tag{5}$$

From (5) we get at once

BOCHNER'S LEMMA. *If* $\Phi_{kk\bullet} + \Phi_k A^{*h}{}_{kh} + \Phi_{k\bullet}A^h{}_{kh} \geqq 0$ *on* M, *then*

$$\Phi_{kk\bullet} + \Phi_k A^{*h}{}_{kh} + \Phi_{k\bullet}A^h{}_{kh} = 0.$$

COROLLARY. *If* $A^h{}_{kh} = 0$, *in particular if the metric is a Kähler metric, then* $\Phi_{kk\bullet} \geqq 0$ *implies* $\Phi_{kk\bullet} = 0$.

15. Pure forms of type $(p, 0)$ and analytic forms. We consider a differential form θ of M which is pure and of type $(p, 0)$

$$\theta = \lambda_{i_1\ldots i_p}\omega^{i_1}\cdots\omega^{i_p}.$$

[6] Hereafter M is supposed compact.

Such forms include the analytic forms since a form is said to be *analytic* if it is pure of type $(p, 0)$ and if, when expressed in local coordinates u^j, u^{*j} of M it depends only on $u^1, \cdots, u^n$, i. e. has analytic functions as components. Writing down the expression for the covariant derivative in local coordinates gives at once the

LEMMA. *θ is analytic if and only if* $\lambda_{i_1 \ldots i_p|k^*} = 0$.

The conditions $d\theta = 0$ and $\delta\theta = 0$ simplify slightly to (1) and (2) respectively

$$\text{(1a)} \qquad \lambda_{i_1 \ldots i_p|k} - \sum_{r=1}^{p} \lambda_{i_1 \ldots k \ldots i_p|i_r} - 2 \sum_{r=1}^{p} \lambda_{i_1 \ldots l \ldots i_p} A^{l}{}_{i_r k} + \sum_{t \neq r}^{p} \sum \lambda_{i_1 \ldots k \ldots l \ldots i_p} A^{l}{}_{i_r i_t} = 0$$

$$\text{(1b)} \qquad \lambda_{i_1 \ldots i_p|l^*} = 0$$

$$\text{(2)} \qquad \lambda_{i_1 \ldots i_{p-1} h|h^*} + 2\lambda_{i_1 \ldots i_{p-1} h} A^{*k}{}_{hk} - \sum_{r=1}^{p-1} \lambda_{i_1 \ldots k \ldots i_{p-1} h} A^{*i_r}{}_{hk} = 0.$$

From which we can draw the immediate conclusions:

THEOREM 1. *If θ is pure of type $(p, 0)$ and $d\theta = 0$, then θ is analytic.*

However the converse is not in general true as it is in the Kähler case, where an analytic form is necessarily harmonic in fact. In the case of non-vanishing torsion we cannot even conclude that a closed analytic form is co-closed (i. e. satisfies $\delta\theta = 0$). However we do have the following theorem for a very special case:

THEOREM 2. *If $\theta = \lambda_i \omega^i$ is an analytic 1-form and if, moreover, $A^i{}_{jk}$ satisfies $A^k{}_{jk} = 0$, then $\delta\theta = 0$ and the form is harmonic if $d\theta = 0$.*

Proof. If θ is analytic, then using the Lemma, (2) reduces to $2\lambda_h A^k{}_{hk}$ which is zero by assumption.

Let us define $\Phi = \lambda^*{}_i \lambda_i$ for $\theta = \lambda_i \omega^i$ an analytic 1-form. Then we have, using $\lambda_{i|k^*} = 0$, $\Phi_{kk^*} = \lambda^*{}_i \lambda_{i|kk^*} + \lambda^*{}_{i|k} \lambda_{i|k}$ and using the interchange formulas $\Phi_{kk^*} = \lambda^*{}_i \lambda_l T_{il^*} + \lambda^*{}_{i|k} \lambda_{i|k}$. Since $\lambda^*{}_{i|k} \lambda_{i|k} \geqq 0$ we get from Bochner's Lemma

THEOREM 3. *If $A^i{}_{jk}$ satisfies $A^k{}_{jk} = 0$ and if the form $\xi^i \xi^{*j} T_{ij^*}$ is positive definite, then there exist no analytic 1-forms on M.*

We may attempt a similar analysis in the case of a pure form of type $(p, 0)$, as above. Define $\Phi = \lambda^*{}_{i_1 \ldots i_p} \lambda_{i_1 \ldots i_p}$. Then,

$$\Phi_{kk} = \lambda^*{}_{i_1 \ldots i_p|k^* k} \lambda_{i_1 \ldots i_p} + \lambda^*{}_{i_1 \ldots i_p} \lambda_{i_1 \ldots i_p|kk^*} + \lambda^*{}_{i_1 \ldots i_p|k} \lambda_{i_1 \ldots i_p|k} + \lambda^*{}_{i_1 \ldots i_p|k^*} \lambda_{i_1 \ldots i_p|k^*}.$$

Assume θ closed, or analytic, so that (1b) prevails, then

$$\Phi_{kk^*} = \lambda^*_{i_1 \ldots i_p} \lambda_{i_1 \ldots i_p | kk^*} + \lambda^*_{i_1 \ldots i_p | k} \lambda_{i_1 \ldots i_p | k}.$$

Using the interchange formula, $\Phi_{kk^*} = \lambda^*_{li_2 \ldots i_p} \lambda_{hi_2 \ldots i_p} T_{lh^*} + \lambda^*_{i_1 \ldots i_p | k} \lambda_{i_1 \ldots i_p | k}$. Again, by Bochner's Lemma we get

THEOREM 4. *If $A^i{}_{jk}$ satisfies $A^k{}_{jk} = 0$ and if $\xi^i \xi^{*j} T_{ij^*}$ is positive definite, then there are no analytic forms on M nor are there any closed pure forms of type $(p, 0)$ on M.*

16. Pure forms of type (1, 1). We consider next a form $\theta = \lambda_{ij^*} \omega^i \omega^{*j}$. The conditions that it be harmonic, i. e. that $d\theta = 0$ and $\delta\theta = 0$ are respectively

(1a) $$\lambda_{ij^*|l} - \lambda_{lj^*|i} - 2\lambda_{kj^*} A^k{}_{il} = 0$$

(1b) $$\lambda_{ij^*|l^*} - \lambda_{il^*|j^*} - 2\lambda_{ik^*} A^{*k}{}_{jl} = 0$$

(1c) $$\lambda_{kj^*|k^*} + 2\lambda_{kj^*} A^{*h}{}_{kh} = 0$$

(1d) $$\lambda_{ik^*|k} + 2\lambda_{ik^*} A^h{}_{kh} = 0.$$

Again we form $\Phi = \lambda_{ij^*} \lambda_{ij^*}$ and compute Φ_{kk^*} for the case where θ is harmonic.

$$\Phi_{kk^*} = \lambda^*_{ij^*|k^*k} \lambda_{ij^*} + \lambda^*_{ij^*} \lambda_{ij^*|kk^*} + \lambda^*_{ij^*|k} \lambda_{ij^*|k} + \lambda^*_{ij^*|k^*} \lambda_{ij^*|k^*}.$$

Let ψ denote the first two terms of this expression, then using $d\theta = 0$, $\delta\theta = 0$ and the interchange formulas and the Bianchi identities we get:

$$\begin{aligned}\psi = &- 2\lambda^*_{ik^*} \lambda_{ij^*} A^{*h}{}_{kh|j^*} - 2\lambda^*_{ij^*} \lambda_{kj^*} A^{*h}{}_{kh|i^*} \\ &+ 2(\lambda^*_{ih^*|k} \lambda_{ij^*} A^h{}_{jk} + \lambda^*_{ij^*} \lambda_{hj^*|k^*} A^h{}_{ik}) \\ &+ \lambda^*_{ik^*} \lambda_{ij^*} R_{jk^*hh^*} - \lambda^*_{hk^*} \lambda_{ij^*} R_{hi^*jk^*} - 2\lambda^*_{ik^*|j^*} \lambda_{ij^*} A^{*h}{}_{kh} - 2\lambda_{ij^*} \lambda_{kj^*|i} A^{*h}{}_{kh} \\ &+ \lambda^*_{ij^*} \lambda_{kj^*} R_{ik^*hh^*} - \lambda^*_{ij^*} \lambda_{hk^*} R_{kj^*ih^*}.\end{aligned}$$

It is clear that $\Phi_{kk^*} \geqq 0$ if $\psi \geqq 0$. In the case $A^h{}_{kh} = 0$ we have

$$\begin{aligned}\psi = &\, 2(\lambda^*_{ih^*|k} \lambda_{ij^*} A^h{}_{jk} + \lambda^*_{ij^*} \lambda_{hj^*|k^*} A^h{}_{ik}) + \lambda^*_{ik^*} \lambda_{ij^*} T_{jk^*} + \lambda^*_{ij^*} \lambda_{kj^*} T_{ik^*} \\ &- \lambda^*_{ij^*} \lambda_{hk^*} (R_{ih^*kj^*} + R_{kj^*ih^*}).\end{aligned}$$

In this case we can state the following weak theorem, which is of some interest since in the Kähler case $\delta_{ij} \omega^i \omega^{*j}$ is a harmonic forms and $\delta_{ij|a} = 0$.

THEOREM 5. *If $A^h{}_{kh} = 0$ and*

$$\lambda^*_{ik^*} \lambda_{ij^*} T_{jk^*} + \lambda^*_{ij^*} \lambda_{kj^*} T_{ik^*} - \lambda^*_{ij^*} \lambda_{hk^*} (R_{ih^*kj^*} + R_{kj^*ih^*})$$

is positive then there are no pure constant harmonic form of type (1, 1).

17. On the non-existence of a harmonic vector field. We consider a differential form of order 1, which we assume self-adjoint and harmonic

$$\theta = \lambda_i \omega^i + \lambda_{i^*} \omega^{*i}, \qquad \lambda^*{}_i = \lambda_{i^*}, \qquad d\theta = 0 = \delta\theta.$$

Any real harmonic vector field on M would give rise to such a form and conversely. The conditions $d\theta = 0$ and $\delta\theta = 0$ give respectively

$$\lambda_{i|k} - \lambda_{k|i} = 2\lambda_j A^j{}_{ik} \tag{1a}$$

$$\lambda_{i|k^*} - \lambda_{k|i^*} = 0 \tag{1b}$$

$$\lambda_{i^*|k^*} - \lambda_{k^*|i^*} = 2\lambda_{j^*} A^{*j}{}_{ik} \tag{1c}$$

$$\lambda_{k|k^*} + A^{*h}{}_{kh}\lambda_k + \lambda_{k^*|k} + A^h{}_{kh^*}\lambda_{k^*} = 0. \tag{2}$$

We let $\Phi = \lambda^*{}_i \lambda_i = \lambda_{i^*}\lambda_i$,

$$\Phi_{kk^*} = \lambda_{i|kk^*}\lambda_i + \lambda_{i^*}\lambda_{i|kk^*} + \lambda_{i^*|k}\lambda_{i|k^*} + \lambda_{i^*|k^*}\lambda_{i|k}.$$

Let ψ denote the first two terms on the right. Then using $d\theta = 0$ and the interchange formulas we get

$$\psi = \lambda_{k|k^*i^*}\lambda_i + 2\lambda_{k|h^*}\lambda_i A^{*h}{}_{ik} + \lambda_{i^*}\lambda_{k|k^*i} + \lambda_{i^*}\lambda_h R_{kh^*ik^*} + 2\lambda_i\lambda_{j|k^*}A^j{}_{ik} + 2\lambda_{i^*}\lambda_j A^j{}_{ik|k^*}.$$

If we assume $A^h{}_{kh} = 0$ we get, using further $d\theta = 0$, $\delta\theta = 0$ and the Bianchi identities

$$\psi = 2\lambda_{k|h^*}\lambda_i A^{*h}{}_{ik} + 2\lambda_{i^*}\lambda_{j|k^*}A^j{}_{ik} + 2\lambda_{i^*}\lambda_j T_{ij^*}.$$

THEOREM 6. *If $A^h{}_{kh} = 0$ and $\xi^j\xi^{*k}T_{jk^*}$ is positive definite, then there are no constant harmonic vector fields on M.*

None of the theorems so far stated involves $R_{ij^*} = R_{hh^*ij^*}$; they use T_{ij^*} which as we have seen in section 10 need not equal R_{ij^*}. Suppose R_{ij^*} to be positive definite, then R_{ij^*} defines a metric tensor on M. It is Hermitian by section 8 (4) and the form $R_{ij^*}\omega^i\omega^{*j} = \Omega^k{}_k$. That this form is closed and hence the metric Kähler was noticed by E. Emry of Zürich in an unpublished thesis. It follows at once from section 8 (6). Thus in this case stronger results known for Kähler manifolds follow.

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY

[1] S. Bochner, "Vector fields and Ricci curvature," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 776-797.

[2] ———, "Curvature and Betti numbers II," *Annals of Mathematics*, vol. 50, No. 1 (1949), pp. 77-93.

[3] S. S. Chern, "Characteristic classes of Hermitian manifolds," *ibid.*, vol. 47, No. 1 (1946), pp. 85-121.

[4] C. Chevalley, *Theory of Lie Groups*, Princeton, New Jersey (1946).

[5] B. Eckmann and H. Guggenheimer, "Formes différentielles et métrique hermitienne sans torsion I, II. Sur les variétés closes à métrique hermitienne sans torsion," *Comptes Rendus de l'Académie des Sciences (Paris)*, vol. 229 (1949), pp. 464, 489, 503.

[6] H. Guggenheimer, "Über komplex-analytische Mannigfaltigkeiten mit Kählerscher Metrik," *Commentarii Mathematici Helvetici*, vol. 25 (1951), pp. 257-297.

[7] H. Hopf, "Zur Topologie der komplexen Mannigfaltigkeiten," *Studies and Essays presented to R. Courant*. New York (1948).

SUR LES MODULES DE REPRÉSENTATION DES ALGÈBRES DE LIE RÉSOLUBLES.*

par J. L. KOSZUL.[1]

1. L'effacement des classes de cohomologie. Soit g une algèbre de Lie sur un corps k et soient M et N deux modules de représentation de g (on dira également deux g-modules). Tout homomorphisme $\phi: M \to N$ définit une application linéaire ϕ' de l'espace $H^*(g, M)$ des classes de cohomologie de g à valeurs dans M dans l'espace $H^*(g, N)$ des classes de cohomologie à valeurs dans N.[2]

Définition. Soit M un g-module de dimension finie sur k. Une classe de cohomologie $c \in H^*(g, M)$ sera dite *effaçable* si il existe un g-module N *de dimension finie* sur k et un *isomorphisme* $\phi: M \to N$ tel que $\phi' \cdot c = 0$.

Ce travail tend essentiellement à déterminer les classes effaçables et à étudier certains modules de représentation en rapport avec cette détermination.

1) *Une classe de cohomologie de degré* 0 *non nulle n'est jamais effaçable.* En effet, une classe $c \in H^0(g, M)$ est un invariant du module M et son image par ϕ' est l'invariant $\phi \cdot c$ du module N.

2) *Toute classe de cohomologie de degré* 1 *est effaçable.* En effet, une classe $c \in H^1(g, M)$ définit une classe d'extension de M par le g-module trivial k. On vérifie immédiatement qu'en prenant pour N une de ces extensions et pour ϕ l'isomorphisme identique de M dans N, on a $\phi' \cdot c = 0$.

3) *Toute classe de cohomologie de degré* 2 *est effaçable.* C'est là un résultat profond, du à Iwasawa, qui exprime en termes de cohomologie l'exis-

* Received November 4, 1953.

[1] Les résultats de cet article ont fait l'objet d'une note "Sur les représentations des algèbres de Lie résolubles," *Comptes Rendus*, 236 (1953), pp. 2371-2372. Il m'a été signalé que le Théorème I du présent travail figure dans la thèse de G. F. Leger (University of Illinois, 1951) avec une démonstration se rattachant au travail de G. Hochschild, "Lie algebras and differentiations in rings of power series," *American Journal of Mathematics*, 62 (1950), pp. 58-80.

[2] Pour les définitions et propriétés fondamentales de la cohomologie des algèbres de Lie, cf. [1] et [2].

tence d'une " splitting algebra " pour toute extension abélienne d'algèbre de Lie.[3]

4) *Pour qu'une algèbre de Lie g sur un corps k de caractéristique* 0 *soit semi-simple, il faut et il suffit que, quel que soit le g-module M de dimension finie sur k, $H^*(g, M)$ ne contienne pas de classe effaçable non nulle.* En effet, si g est semi-simple, pour tout isomorphisme ϕ de M dans un g-module N de dimension finie, il existe un homomorphisme ψ de N sur M tel que $\psi\phi$ soit l'identité de M. Il en résulte que ϕ' est un isomorphisme. Inversement, si pour tout g-module M de dimension finie $H^1(g, M)$ ne contient pas de classe effaçable autre que 0, ceci signifie d'après 2) que $H^1(g, M) = (0)$, donc que tous les g-modules de dimension finie sont complèment réductibles.

5) *Soit g une algèbre de Lie sur un corps k de caractéristique* 0. *Si les classes de cohomologie de degré* 3 *de g à valeurs dans k sont effaçables, alors g est résoluble.* Soit en effet $g = b + r$ une décomposition de Levi de g (r étant le radical de g et b une sous-algèbre semi-simple supplémentaire de r dans g). Tout g-module L peut être considéré comme b-module et il existe un homomorphisme canonique du complexe des cochaînes de g à valeurs dans L sur le complexe des cochaînes de b à valeurs dans L. Cet homomorphisme définit pour tout entier $p \geqq 0$ une application ρ_L de $H^p(g, L)$ *sur* $H^p(b, L)$.[4] Si $b \neq (0)$, alors $H^3(b, k) \neq (0)$ et il existe donc une classe $c \in H^3(g, k)$ telle que $\rho_k \cdot c \neq 0$. Si il existe un isomorphisme ϕ de k dans un g-module N qui réalise l'effacement de c, cet isomorphisme étant compatible avec les structures de b-modules va définir une application ϕ'_b: $H^3(b, k) \to H^3(b, N)$ qui vérifie la relation $\phi'_b\rho_k = \rho_N\phi'$. Par conséquent, $\rho_k \cdot c$ serait une classe effaçable non nulle de $H^3(b, k)$, contrairement à 4). On a donc $b = (0)$ et g est résoluble.

Le Théorème suivant est l'un des principaux résultats que l'on a en vue:

THÉORÈME I. *Si g est une algèbre de Lie résoluble sur un corps de caractéristique* 0, *alors, quel que soit le g-module M de dimension finie, toutes les classes de $H^p(g, M)$ sont effaçables pour $p > 0$.*

Sa démonstration sera donnée à la fin du paragraphe 8. Compte tenu de 5), on voit que l'on obtiendra une manière de *caractériser* les algèbres de Lie résolubles par les propriétés d'effacement de leurs modules de représentation.

---

[3] Cf. [3].

[4] Ceci résulte de l'existence d'un homomorphisme de g sur b inverse à gauche de l'isomorphisme $b \to g$.

OHIO UNIVERSITY
LIBRARY

2. Le module $V(g)$ et son interprétation. Soit A une algèbre associative avec unité de dimension *infinie* sur un corps k. Etant donné un module à gauche M sur A, on dira qu'un élément de M est de *caractère fini* si le sous-module qu'il engendre est de dimension finie sur k. Pour que $u \in M$ soit de caractère fini, il faut et il suffit que son annulateur (idéal à gauche des $a \in A$ tels que $au = 0$) soit de codimension finie dans A. Il revient au même de dire que son annulateur contient un idéal *bilatère* de codimension finie. Les éléments de caractère fini de M constituent un sous-module. Si h est un homomorphisme de M dans un A-module à gauche N, l'image par h d'un élément de caractère fini est un élément de caractère fini de N. On définit de manière analogue les éléments de caractère fini pour un module à droite sur A.

Soit $E(g)$ l'algèbre enveloppante de l'algèbre de Lie g. Les g-modules ne sont autres que les modules unitaires à gauche sur l'algèbre associative $E(g)$. Pour tout g-module M, on définit une structure *canonique* de g-module dans l'espace $\mathfrak{L}(E(g), M)$ des applications k-linéaires de $E(g)$ dans M en posant $(xf)\cdot e = f\cdot(ex)$ pour $f \in \mathfrak{L}(E(g), M)$, $e \in E(g)$ et $x \in g$. L'isomorphisme $\phi_M : M \to \mathfrak{L}(E(g), M)$ défini par $(\phi_M \cdot a)\cdot e = ea$ pour $a \in M$ et $e \in E(g)$ est alors compatible avec les structures de g-modules. On démontre que, pour tout g-module M et tout entier $p > 0$, $H^p(g, \mathfrak{L}(E(g), M) = (0)$.[5] Toute classe de cohomologie dans $H^p(g, M)$ est donc annulée par ϕ_M lorsque $p > 0$. On observera que, pour $M \neq (0)$, $\mathfrak{L}(E(g), M)$ est *de dimension infinie* sur k, si bien que ϕ_M ne réalise pas l'effacement des classes de $H^p(g, M)$. On va cependant voir que cet isomorphisme ϕ_M est utile dans le problème que l'on a en vue.

Notons $\mathfrak{L}_f(E(g), M)$ le sous-module des éléments de caractère fini du g-module $\mathfrak{L}(E(g), M)$ considéré comme module à gauche sur $E(g)$; c'est le g-module des applications k-linéaires de $E(g)$ dans M qui sont nulles sur un idéal bilatère de codimension finie dans $E(g)$. Si M est de dimension finie sur k, on a $\phi_M \cdot M \subset \mathfrak{L}_f(E(g), M)$. On notera encore ϕ'_M l'application des espaces de cohomologie associée à l'isomorphisme $\phi_M : M \to \mathfrak{L}_f(E(g), M)$.

THÉORÈME 2. *Quel que soit le g-module de dimension finie M, les classes effaçables de $H^*(g, M)$ sont les zéros de l'application*

$$\phi'_M : H^*(g, M) \to H^*(g, \mathfrak{L}_f(E(g), M)).$$

---

[5] Ce résultat est lié au suivant: il existe un complexe acyclique pour les degrés $\neq 0$ et muni d'une structure de module libre sur $E(g)$ tel que, pour tout g-module M, $\mathrm{Hom}_{E(g)}(C, M)$ soit le complexe des cochaînes de g à valeurs dans M. On en déduit que le complexe $\mathrm{Hom}_{E(g)}(C, \mathfrak{L}(E(g), M))$ qui s'identifie au complexe des applications

Supposons en effet qu'une classe $c \in H^p(g, M)$ soit un zéro de ϕ'_M. Soit f un cocycle de g à valeurs dans M appartenant à la classe c. Le cocycle $\phi_M f$ de g à valeurs dans $\mathfrak{L}_f(E(g), M)$ est un cobord, donc de la forme dg où g est une cochaîne de degré $p-1$ de g à valeurs dans $\mathfrak{L}_f(E(g), M)$. Soit N le sous-module de $\mathfrak{L}_f(E(g), M)$ engendré par les valeurs de g et l'image de ϕ_M. Etant engendrable par un nombre fini d'éléments de caractère fini, N est de dimension finie sur k.[6] L'isomorphisme $\phi_M : M \to N$ réalise visiblement l'effacement de c.

Supposons inversement que c soit une classe effaçable de $H^*(g, M)$ et que ψ soit un isomorphisme de M dans un g-module N qui en réalise l'effacement. Soit ρ une application k-linéaire de N sur M telle que $\rho\psi$ soit l'identité de M. Les applications ψ et ρ définissent des homomorphismes de g-modules $\psi_e : \mathfrak{L}_f(E(g), M) \to \mathfrak{L}_f(E(g), N)$ et $\rho_e : \mathfrak{L}_f(E(g), N) \to \mathfrak{L}_f(E(g), M)$ tels que $\rho_e\psi_e$ soit l'identité de $\mathfrak{L}_f(E(g), M)$. Par suite ψ'_e est un isomorphisme de $H^*(g, \mathfrak{L}_f(E(g), M))$ dans $H^*(g, \mathfrak{L}_f(E(g), N))$. On a d'autre part $\psi_e\phi_M = \phi_N\psi$, par suite $\psi'_e\phi'_M \cdot c = \phi'_N\psi' \cdot c = 0$ et $\phi'_M \cdot c = 0$.

On est ainsi conduit à étudier la cohomologie de g à valeurs dans le g-module $\mathfrak{L}_f(E(g), M)$. Celle ci se ramène à la cohomologie de g à valeurs dans $\mathfrak{L}_f(E(g), k)$ muni de la structure canonique de g-module. En effet, M étant supposé de dimension finie sur k, on a un isomorphisme naturel du produit tensoriel (sur k) $\mathfrak{L}(E(g), k) \otimes M$ sur $\mathfrak{L}(E(g), M)$, qui est compatible avec les structures de g-modules si dans le produit tensoriel on définit $x(f \otimes u) = xf \otimes u$ pour $x \in g$, $f \in \mathfrak{L}(E(g), k)$ et $u \in M$. Cet isomorphisme applique donc $\mathfrak{L}_f(E(g), k) \otimes M$ sur $\mathfrak{L}_f(E(g), M)$ et il existe un isomorphisme naturel de $H^p(g, \mathfrak{L}_f(E(g), k)) \otimes M$ sur $H^p(g, \mathfrak{L}_f(E(g), M))$ pour tout p. *Le module* $\mathfrak{L}_f(E(g), k)$ *sera noté* $V(g)$.

Soit $A(g)$ l'idéal bilatère de l'algèbre enveloppante de g qui est engendré par les éléments de g.[7] Sa codimension est 1 et tous les idéaux $A(g)^r$ (r entier > 0) sont de codimension finie. Etant donné un g-module M, les éléments de M dont l'annulateur contient un idéal de la forme $A(g)^r$ seront appelés des éléments de *caractère nilpotent* de M. Pour qu'un élément u de M soit de caractère nilpotent, il faut et il suffit qu'il soit de caractère fini et que la représentation de g dans M induise dans le sous-module $E(g)u$ une représen-

---

k-linéaires de C dans M est acyclique pour les degrés $\neq 0$. Les démonstrations de ces résultats, qui ne seront pas utilisés ici, figureront dans le livre de H. Cartan et S. Eilenberg (*Princeton Mathematical Series*).

[6] Une cochaîne de g à valeurs dans M étant une application multilinéaire alternée de g dans M, ses valeurs constituent un sous-espace de dimension finie dans M.

[7] On identifiera toujours une algèbre de Lie avec le sous-espace de son algèbre enveloppante qui lui correspond.

tation par des endomorphismes nilpotents. On aura accessoirement à utiliser le sous-module des éléments de caractère nilpotent de $\mathfrak{L}(E(g), k)$. *Il sera noté* $W(g)$; c'est le module des fonctions k-linéaires sur $E(g)$ nulles sur un idéal de la forme $A(g)^r$.

On donnera dans ce qui suit une interprétation des cochaînes de g à valeurs dans les modules $V(g)$ et $W(g)$ lorsque g est l'algèbre de Lie d'un groupe de Lie G, le corps de base étant alors le corps des réels R. D'une manière précise, on supposera g identifiée à l'algèbre de Lie des champs de vecteurs invariants par les translations à droite sur G. Soit L l'algèbre des fonctions différentiables de classe infinie à valeurs réelles sur G. Soit $\theta(x)$ la transformation infinitésimale du groupe des translations à gauche associée au champ x. En possant $xF = \theta(x) \cdot F$ pour tout $x \in g$ et $F \in L$, on définit dans L une structure de module à gauche sur $E(g)$. Soit x' le champ de vecteurs invariant à gauche qui coïncide avec x au point neutre ϵ de G et soit $\theta(x')$ la transformation infinitésimale du groupe des translations à droite qui correspond à x'. On a $\theta(x')\theta(y') - \theta(y')\theta(x') = -\theta([x, y]')$ et par suite, on définit dans L une structure de module à *droite* sur $E(g)$ par la condition $Fx = \theta(x') \cdot F$ pour $x \in g$ et $F \in L$. On observera que $(eF)e' = e(Fe')$ quels que soient $e, e' \in E(g)$, $F \in L$ et que d'autre part, les fonctions xF et Fx ont même valeur au point ϵ.

Pour tout $F \in L$, on notera $\lambda \cdot F$ la fonction R-linéaire définie sur $E(g)$ par $(\lambda \cdot F) \cdot e = (eF)(\epsilon)$. On a $(\lambda \cdot (xF)) \cdot e = (exF)(\epsilon) = (\lambda \cdot F) \cdot (ex)$. L'application $\lambda : L \to \mathfrak{L}(E(g), R)$ est donc compatible avec les structures de module à *gauche* sur $E(g)$. On définit dans $\mathfrak{L}(E(g), R)$ une structure de module à droite sur $E(g)$ commutant avec la structure à gauche envisagée jusqu'ici en posant $(he) \cdot e' = h \cdot (ee')$ quels que soient $h \in \mathfrak{L}(E(g), R)$, $e, e' \in E(g)$. On a $(\lambda \cdot (Fx)) \cdot e = (e(Fx))(\epsilon) = ((eF)x)(\epsilon) = (xeF)(\epsilon) = (\lambda \cdot F) \cdot xe$; par suite, l'application λ est également compatible avec les structures de modules à droite sur $E(g)$. Le noyau de λ est l'idéal des fonctions qui sont nulles ainsi que toutes leurs dérivées au point ϵ.

Soit $L_f \subset L$ la sous-algèbre des fonctions sur G dont les translatées à gauche (ou, ce qui revient au même, les translatées à droite) sont dans un espace vectoriel de dimension finie sur R. Ces fonctions étant analytiques et de caractère fini pour la structure de module à gauche sur $E(g)$ définie dans L, λ *applique isomorphiquement* L_f *dans* $V(g)$. Soit $L_n \subset L_f$ la sous-algèbre des éléments de L_f qui sont de caractère nilpotent pour la structure de module à gauche sur $E(g)$, ou, ce qui revient au même, dont l'image par λ est dans $W(g)$. Il est clair que $W(g)$ est également le sous-module des éléments de $\mathfrak{L}(E(g), R)$ qui sont de caractère nilpotent pour la structure de module à

droite sur $E(g)$. Les éléments de L_n peuvent donc aussi se définir comme étant les éléments de L_f qui sont de caractère nilpotent pour la structure de module à droite sur $E(g)$.

Supposons maintenant G *simplement connexe.* Tout élément $h \in V(g)$ engendre un sous-module $E(g)h \subset V(g)$ de dimension finie sur R. Puisque G est simplement connexe, à la représentation linéaire de g dans $E(g)h$ qui correspond à la structure de module à gauche sur $E(g)$ est associée une représentation linéaire $s \to r(s)$ du groupe G dans l'espace vectoriel $E(g)h$.[8] La fonction $F(s)$ définie sur G par $F(s) = (r(s)\cdot h)\cdot I$, où I désigne l'unité de $E(g)$, est un élément de L_f dont l'image par λ est précisément h. Dans ce cas donc, λ applique isomorphiquement L_f sur $V(g)$ et L_n sur $W(g)$.

THÉORÈME 3. *Soit g l'algèbre de Lie d'un groupe simplement connexe G. Le complexe des cochaînes de g à valeurs dans $V(g)$ est canoniquement isomorphe au complexe des formes différentielles sur G dont les translatées sont dans un espace vectoriel de dimension finie. Dans cet isomorphisme, les cochaînes de g à valeurs dans $W(g)$ correspondent aux formes différentielles sur G dont les translatées sont dans un espace vectoriel de dimension finie où les transformations infinitésimales du groupe des translations à droite sont nilpotentes.*

Soit $\omega(C_1, C_2, \cdots C_p)$ une fonction L-linéaire alternée, à valeurs dans L, de p champs de vecteurs de classe infinie sur G, autrement dit, une forme differentielle de degré p sur G. Si les translatées à droite (ou ce qui revient au même, à gauche) de cette forme sont dans un espace vectoriel de dimension finie, alors, pour des C_i invariants à droite, $\omega(C_1, C_2, \cdots, C_p)$ sera dans L_f. Par restriction des formes aux champs invariants à droite, c'est à dire à g, on définit donc une application linéaire canonique λ_c de l'espace des formes différentielles dont les translatées sont dans un espace de dimension finie dans l'espace des cochaînes de g à valeurs dans L_f qui s'identifie à $V(g)$ par l'isomorphisme λ. Le module des champs de vecteurs sur G étant un L-module libre admettant pour base une base de g sur R, λ_c est un isomorphisme sur. D'autre part, la différentielle d'une forme $\omega(C_1, C_2, \cdots, C_p)$ peut s'exprimer par la formule

$$(d\cdot\omega)(C_1, C_2, \cdots C_{p+1}) = \sum_{i<j} (-1)^{i+j+1}\omega([C_i, C_j], \cdots \hat{C}_i, \cdots \hat{C}_j, \cdots)$$
$$+ \sum_i (-1)^i \theta(C_i)\cdot\omega(C_1, C_2, \cdots \hat{C}_i, \cdots C_{p+1})$$

où $\theta(C_i)$ est la transformation infinitésimale correspondant au champ C_i.

[8] Cf. Chevalley, " Theory of Lie Groups," *Princeton*, p. 113.

Comme cette formule est aussi celle qui donne le cobord d'une cochaîne de g, l'isomorphisme λ_e est compatible avec les structures de complexes. Il fait correspondre aux formes qui sont de caractère nilpotent vis à vis des transformations infinitésimales du groupe des translations à droite [9] les cochaînes de g à valeurs dans $W(g)$.

3. Remarques préliminaires. *Invariants des modules $V(g)$ et $W(g)$.* Pour qu'une fonction k-linéaire f sur $E(g)$ soit un invariant du g-module $\mathfrak{L}(E(g), k)$, il faut et il suffit qu'elle soit nulle sur l'idéal $A(g) \subset E(g)$ Soit β l'homomorphisme de l'algèbre $E(g)$ sur k ayant pour noyau $A(g)$. En associant à tout $c \in k$ la fonction $e \to c(\beta \cdot e)$ de $E(g)$ dans k, on obtient un isomorphisme de k sur le sous-module des invariants de $\mathfrak{L}(E(g), k)$ qui est contenu dans $W(g) \subset V(g)$. Il lui correspond des applications naturelles $H^*(g, k) \to H^*(g, W(g)) \to H^*(g, V(g))$. Pour le degré 0, ce sont des isomorphismes de $k = H^0(g, k)$ sur $H^0(g, W(g))$ et sur $H^0(g, V(g))$.[10]

Cas d'un produit direct d'algèbres de Lie. Soient g_1 et g_2 deux algèbres de Lie sur un corps k. On identifiera l'algèbre $E(g_1 \times g_2)$ au produit tensoriel (sur k) $E(g_1) \otimes E(g_2)$ au moyen de l'isomorphisme qui prolonge l'application $(x, y) \to x \otimes I + I \otimes y$ de $g_1 \times g_2$ dans $E(g_1) \otimes E(g_2)$ (les unités de $E(g_1)$ et $E(g_2)$ étant notées I). Soit π l'isomorphisme de $\mathfrak{L}(E(g_1), k) \otimes \mathfrak{L}(E(g_2), k)$ dans $\mathfrak{L}(E(g_1 \times g_2), k)$ qui associe à $f_1 \otimes f_2$ la fonction $e_1 \otimes e_2 \to (f_1 \cdot e_1)(f_2 \cdot e_2)$ pour $e_1 \in E(g_1)$ et $e_2 \in E(g_2)$. Les espaces $\mathfrak{L}(E(g_1), k)$ et $\mathfrak{L}(E(g_2), k)$ étant des modules à gauche sur $E(g_1)$ et $E(g_2)$ respectivement, leur produit tensoriel est muni d'une structure de module à gauche sur $E(g_1) \otimes E(g_2)$. L'isomorphisme π est un homomorphisme de modules à gauche sur $E(g_1) \otimes E(g_2)$ et applique donc $V(g_1) \otimes V(g_2)$ dans $V(g_1 \times g_2)$. Quoique l'image de π ne soit qu'un *sous-module* de $\mathfrak{L}(E(g_1 \times g_2), k)$, on vérifie que $\pi \cdot (V(g_1) \otimes V(g_2)) = V(g_1 \times g_2)$. Cet isomorphisme π définit donc un isomorphisme canonique du produit tensoriel du complexe des cochaînes de g_1 à valeurs dans $V(g_1)$ et du complexe des cochaînes de g_2 à valeurs dans $V(g_2)$ sur le complexe des cochaînes de $g_1 \times g_2$ à valeurs dans $V(g_1 \times g_2)$. On en déduit le

[9] En général, l'algèbre des formes de caractère nilpotent vis à vis des translations à droite est distincte de celle des formes de caractère nilpotent vis à vis des translations à gauche. La structure multiplicative dans $V(g)$ et $W(g)$ qui résulte de l'isomorphisme λ pourrait se définir algébriquement dans le cas général en prenant l'application transposée de l'isomorphisme diagonal de $E(g)$ dans $E(g) \otimes E(g)$.

[10] Si g est une algèbre semi-simple sur un corps de caractéristique 0, l'isomorphisme de k dans $W(g) \subset V(g)$ induit un isomorphisme de $H^p(g, k)$ sur $H^p(g, V(g))$ et $H^p(g, W(g))$ *pour tout entier* p.

LEMME 1. *Si g_1 et g_2 sont des algèbres de Lie sur un même corps k il existe un isomorphisme canonique du produit tensoriel (sur k)*

$$H^*(g_1, V(g_1)) \otimes H^*(g_2, V(g_2)) \text{ sur } H^*(g_1 \times g_2, V(g_1 \times g_2)).$$

4. Cas des algèbres de Lie abéliennes.

THÉORÈME 4. *Si g est une algèbre de Lie abélienne, alors $H^p(g, W(g)) = H^p(g, V(g)) = (0)$ pour tout entier $p > 0$.*

Soit (x_i) $(i = 1, 2, \cdots n)$ une base de g et soit (x'_i) la base duale dans le dual de g. L'espace $W(g)$ s'identifie à l'algèbre des polynômes en les x'_i, la structure de g-module dans $W(g)$ étant celle qui associe à chaque x_i la dérivation par rapport à x_i dans $W(g)$. On sait que, dans ces conditions, $H^p(g, W(g)) = (0)$ pour tout $p > 0$.[11] Pour démontrer le second résultat, il suffit, en vertu du Lemme 1, de l'établir dans le cas où g est de dimension 1 sur k, auquel cas l'énoncé se ramène à $H^1(g, V(g)) = (0)$.

Soit $f(x)$ un cocycle de degré 1 de g à valeurs dans $V(g)$ et soit u un élément non nul de g. Puisque $f(u) \in V(g)$, c'est une fonction linéaire sur l'algèbre $E(g)$ des polynômes en u qui est nulle sur un idéal de codimension finie, donc sur un idéal engendré par un polynôme $P(u) \neq 0$. Soit h la fonction linéaire sur $E(g)$ définie par $h \cdot I = 0$ et $h \cdot u^p = f(u) \cdot u^{p-1}$ pour $p > 0$. Elle est nulle sur l'idéal engendré par $uP(u)$. C'est donc un élément de $V(g)$ ou encore, une cochaîne de degré 0 de g à valeurs dans $V(g)$. Son cobord dh est un cocycle de degré 1 associant à tout $x = cu$ une fonction $(dh)(x)$ sur $E(g)$. On a $((dh)(x)) \cdot u^p = (xh) \cdot u^p = h \cdot (u^p x) = cf(u) \cdot u^p = f(x) \cdot u^p$. Par conséquent, $dh = f$, ce qui achève la démonstration.

On donnera dans ce qui suit une généralisation du Théorème 4 qui sera utilisée aux paragraphes 7 et 8.

Soient g une algèbre de Lie quelconque et M un g-module. On notera α_M l'application linéaire $f \to f \cdot I$ de $\mathfrak{L}(E(g), M)$ sur M, I désignant l'unité de $E(g)$. On définit deux représentations linéaires $x \to \rho_1(x)$ et $x \to \rho_2(x)$ de g dans l'espace $\mathfrak{L}(E(g), M)$ en posant

$$(\rho_1(x) \cdot f) \cdot e = x(f \cdot e) + f \cdot (ex), \qquad (\rho_2(x) \cdot f) \cdot e = f \cdot (ex)$$

quels que soient $f \in \mathfrak{L}(E(g), M)$, $e \in E(g)$ et $x \in g$. La seconde représentation est celle qui correspond à la structure de g-module envisagée au paragraphe 1. Ces représentations se prolongent en des homomorphismes $e \to \rho_1(e)$ et $e \to \rho_2(e)$ de l'algèbre $E(g)$ dans l'algèbre des endomorphismes

[11] Le raisonnement classique, par récurrence sur l'entier n, sera généralisé dans le Théorème 6.

de l'espace $\mathfrak{L}(E(g), M)$. On va montrer que ces représentations sont équivalentes.

LEMME 2. *Il existe un automorphisme θ et un seul de l'espace $\mathfrak{L}(E(g), M)$ tel que*

1° $\theta\rho_1(x) = \rho_2(x)\theta$ *pour tout* $x \in g$,

2° $\alpha_M\theta = \alpha_M$.

Un endomorphisme θ de l'espace $\mathfrak{L}(E(g), M)$ vérifiant ces conditions est nécessairement donné par

$$(\theta\cdot f)\cdot e = (\theta\cdot f)\cdot(Ie) = (\rho_2(e)\theta\cdot f)\cdot I = (\theta\rho_1(e)\cdot f)\cdot I = (\rho_1(e)\cdot f)\cdot I$$

lorsque $e \in E(g)$ et $f \in \mathfrak{L}(E(g), M)$. Réciproquement,, l'endomorphisme θ défini par $(\theta\cdot f)\cdot e = (\rho_1(e)\cdot f)\cdot I$ vérifie les conditions du Lemme. Il reste à démontrer que θ possède un inverse. Soit μ l'isomorphisme de l'algèbre $E(g)$ dans l'algèbre $E(g) \otimes E(g)$ qui est défini par les conditions $\mu\cdot I = I \otimes I$ et $\mu\cdot x = x \otimes I + I \otimes x$ pour tout $x \in g$.[12] On notera $e \to \bar{e}$ l'automorphisme involutif de l'espace $E(g)$ qui conserve l'unité et transforme tout monôme $x_1 x_2 \cdot \cdot \cdot x_p$ $(x_i \in g)$ en le monôme $(-1)^p x_p x_{p-1} \cdot \cdot \cdot x_1$. Soit ψ l'endomorphisme de l'espace $\mathfrak{L}(E(g), M)$ défini par $(\psi\cdot f)\cdot e = \sum_i \bar{e}_i(f\cdot e'_i)$ lorsque $f \in \mathfrak{L}(E(g), M)$ et $\mu\cdot e = \sum_i e_i \otimes e'_i$. On a $\mu\cdot(ex) = \sum_i (e_i x) \otimes e'_i + \sum_i e_i \otimes (e'_i x)$. Comme $\overline{e_i x} = -x\bar{e}_i$, il en résulte que, pour tout $x \in g$, on a $\rho_1(x)\psi = \psi\rho_2(x)$. D'autre part, $\alpha_M\psi = \alpha_M$ et par suite $\alpha_M\psi\theta = \alpha_M\theta\psi = \alpha_M$. Si $x \in g$ et $e \in E(g)$, on a donc

$$(\psi\theta\cdot f)\cdot(ex) = (\rho_1(x)\psi\theta\cdot f)\cdot e - x((\psi\theta\cdot f)\cdot e) = (\psi\theta\rho_1(x)\cdot f)\cdot e - x((\psi\theta\cdot f)\cdot e).$$

Il en résulte que l'ensemble des éléments $e \in E(g)$ tels que $(\psi\theta\cdot f)\cdot e = f\cdot e$ pour tout $f \in \mathfrak{L}(E(g), M)$ constitue un idéal à droite de $E(g)$. Puisqu'il contient I, c'est donc que $\psi\theta$ est l'identité. La relation $\rho_2(x)\theta\psi = \theta\psi\rho_2(x)$ montre d'autre part que $\theta\psi$ est l'identité. Par suite ψ est l'inverse de θ.

LEMME 3. *Le sous-espace des invariants de $\mathfrak{L}(E(g), M)$ pour la représentation ρ_1 est appliqué isomorphiquement sur M par l'application α_M.*

Soit en effet f un invariant de $\mathfrak{L}(E(g), M)$ pour la représentation ρ_1; $\theta\cdot f$ est un invariant de $\mathfrak{L}(E(g), M)$ pour la représentation ρ_2. C'est donc une fonction linéaire de $E(g)$ dans M nulle sur l'idéal $A(g)$. Si $\alpha_M\cdot f = 0$, alors $(\theta\cdot f)\cdot I = \alpha_M\theta\cdot f = \alpha_M\cdot f = 0$ et par suite $\theta\cdot f = 0$, donc $f = 0$. D'autre part, quel que soit $a \in M$, il existe une fonction $f \in \mathfrak{L}(E(g), M)$ et une seule

[12] Cf. (7).

telle que $f \cdot I = a$ et $f \cdot A(g) = (0)$. Cette fonction est un invariant de $\mathfrak{L}(E(g), M)$ pour la représentation ρ_2. Par suite $\theta^{-1} \cdot f$ est un invariant de $\mathfrak{L}(E(g), M)$ pour la représentation ρ_1 et l'on a $\alpha_M \theta^{-1} \cdot f = \alpha_M \cdot f = a$.

LEMME 4. *Si g est abélienne et si M est de dimension finie, l'espace $V(g) \otimes M$ identifié à un sous-espace de $\mathfrak{L}(E(g), M)$ est stable par θ et θ^{-1}. L'application α_M induit un isomorphisme de l'espace des invariants de $V(g) \otimes M$ pour la représentation ρ_1 sur l'espace M. Si de plus la représentation de g dans M est une représentation par endomorphismes nilpotents, ces résultats subsistent en remplaçant $V(g)$ par $W(g)$.*

Désignons par (ρ_1), (ρ_2) et (ρ_3) les trois structures de g-modules définies dans $L = \mathfrak{L}(E(g), M)$ par les représentations $x \to \rho_1(x)$, $x \to \rho_2(x)$ et $x \to \rho_3(x) = \rho_1(x) - \rho_2(x)$. Si g est abélienne, les endomorphismes $\rho_1(x)$, $\rho_2(y)$ et $\rho_3(z)$ commutent entre eux quels que soient $x, y, z \in g$. Il en résulte que, si un élément de L est de caractère fini pour deux de ces structures, il l'est également pour la troisième. Si M est de dimension finie, tout élément de L est de caractère fini pour (ρ_3) et les éléments de caractère fini pour (ρ_1) sont donc aussi les éléments de caractère fini pour (ρ_2). Puisque $\theta \rho_1(x) = \rho_2(x) \theta$ pour tout $x \in g$, le sous-espace $V(g) \otimes M$ qui est le sous-espace des éléments de caractère fini pour (ρ_2) ou (ρ_1) est donc stable par θ. De même, la relation $\theta^{-1} \rho_2(x) = \rho_1(x) \theta^{-1}$ montre qu'il est stable par θ^{-1}. Tout élément de L invariant pour (ρ_1) étant de caractère fini pour cette structure est donc dans $V(g) \otimes M$. La seconde partie de l'énoncé résulte ainsi du Lemme 3. Dans le cas où la représentation de g dans M est une représentation par endomorphismes nilpotents, on obtiendra des résultats analogues pour $W(g) \otimes M$ en remarquant que les éléments de L sont tous de caractère nilpotent pour la structure (ρ_3) et que les invariants de L pour la structure (ρ_1) sont de caractère nilpotent pour cette structure.

On utilisera dans la suite la situation suivante : b est un idéal de l'algèbre de Lie g sur k ; U est un g-module arbitraire ; λ est un homomorphisme de U, considéré comme b-module, dans un b-module X. En composant l'homomorphisme canonique du complexe $C(g, U)$ des cochaînes de g à valeurs dans U dans le complexe $C(b, U)$ des cochaînes de b à valeurs dans U avec l'homomorphisme de $C(b, U)$ dans $C(b, X)$ qui prolonge λ, on obtient un homomorphisme $\sigma : C(g, U) \to C(b, X)$ induisant une application $\sigma^* : H^*(g, U) \to H^*(b, X)$. On utilisera dans $H^*(b, U)$ la structure de g/b-module défini comme suit. Dans le complexe $C(b, U)$, on définit une structure de g-module en posant

$$(xf)(y_1, y_2, \cdots y_p) = x(f(y_1, y_2, \cdots y_p)) - \sum_{i=1}^{i=p} f(y_1, \cdots [x, y_i], \cdots y_p)$$

pour f de degré p dans $C(b, U)$, $x \in g$ et $y_1, y_2, \cdots y_p \in b$. Comme $x(d \cdot f) = d \cdot (xf)$ pour tout $f \in C(b, U)$ et $x \in g$, cette structure se retrouve dans $H^*(b, U)$ où elle induit une structure de g/b-module car $xc = 0$ pour tout $c \in H^*(b, U)$ lorsque $x \in b$.

LEMME 5. *Pour que σ^* soit un isomorphisme de $H^*(g, U)$ sur $H^*(b, X)$, il suffit que les conditions suivantes soient vérifiées:*

1° $H^j(g/b, H^*(b, U)) = (0)$ *pour tout* $j > 0$,

2° *l'application* $\lambda' : H^*(b, U) \to H^*(b, X)$ *applique isomorphiquement le sous-espace des invariants du g/b-module $H^*(b, U)$ sur $H^*(b, X)$.*

Ce Lemme est une conséquence des résultats sur les suites spectrales relatives à un idéal établis par Hochschild et Serre.[13] Soit A^j le sous-complexe de $C(g, U)$ dont les éléments de degré p sont les cochaînes de g à valeurs dans U qui sont nulles lorsque $p - j + 1$ variables au moins sont dans b. A la filtration de $C(g, U)$ par les A^j est associée une suite spectrale E_s $(s = 0, 1, 2, \cdots)$. D'après le Théorème 4 de [2], il existe pour tout entier $j \geqq 0$ un isomorphisme canonique de l'espace $E^j{}_2$ des éléments de E_2 ayant j pour degré de filtration sur l'espace $H^j(g/b, H^*(b, U))$, la structure de g/b-module de $H^*(b, U)$ étant celle qui a été décrite plus haut. Le sous-complexe A^1 est le noyau de l'homomorphisme $C(g, U) \to C(b, U)$. Cet homomorphisme induit donc une application de E_2 dans $H^*(b, U)$ qui, composée avec λ', donne une application canonique $\sigma^*{}_2 : E_2 \to H^*(b, X)$. Cette application est nulle sur $E^j{}_2$ pour $j > 0$. On vérifie d'autre part que, après identification de $E^0{}_2$ à $H^0(g/b, H^*(b, U))$, c'est à dire à l'espace des invariants du g/b-module $H^*(b, U)$, $\sigma^*{}_2$ coïncide sur $E^0{}_2$ avec la restriction de λ'. Les hypothèses du Lemme signifient donc que $\sigma^*{}_2$ applique isomorphiquement E_2 sur $H^*(b, X)$. Ceci entraîne que σ^* applique isomorphiquement $H^*(g, U)$ sur $H^*(b, X)$ d'après un résultat bien connu de la technique des suites spectrales.

On particularisera les hypothèses de ce Lemme sous deux formes. Pour commencer, on supposera $U = V(g/b)$, $X = k$ et λ sera l'homomorphisme canonique $\alpha : V(g/b) \to k$ (les structures de b-modules étant celles qui correspondent aux représentations nulles). On notera dans ce cas σ_v l'homomorphisme de $C(g, V(g/b))$ dans $C(b, k)$. On envisagera également le cas où $U = W(g/b)$, $X = k$, λ étant alors la restriction de α à $W(g/b)$. Il conduit à un homomorphisme $\sigma_w : C(g, W(g/b)) \to C(b, k)$.

[13] Cf. [2], paragraphes 2 et 3.

THÉORÈME 5. *Si b est un idéal de l'algèbre de Lie g sur k tel que g/b soit abélienne, alors σ^*_v est un isomorphisme de $H^*(g, V(g/b))$ sur $H^*(b, k)$. Si de plus g est une algèbre de Lie nilpotente, alors σ^*_w est un isomorphisme de $H^*(g, W(g/b))$ sur $H^*(b, k)$.*

Puisque la représentation de b dans $V(g/b)$ est la représentation nulle, le g/b-module $H^*(b, V(g/b))$ s'identifie au produit tensoriel $V(g/b) \otimes H^*(b, k)$ et l'application $\lambda' : H^*(b, V(g/b)) \to H^*(b, k)$ devient l'application $v \otimes a \to (\alpha \cdot v)a$ pour $v \in V(g/b)$ et $a \in H^*(b, k)$. Puisque g/b est abélienne et que $H^*(b, k)$ est de dimension finie, la condition 2° du Lemme 5 est donc vérifiée en vertu du Lemme 4. Il résulte également du Lemme 4 que les représentations de g/b dans $V(g/b) \otimes H^*(b, k)$ définies par $\rho_1(x)(v \otimes a) = xv \otimes a + v \otimes xa$ et $\rho_2(x) \cdot (v \otimes a) = xv \otimes a$ sont équivalentes. Si l'on calcule $H^j(g/b, V(g/b) \otimes H^*(b, k))$ en prenant dans $V(g/b) \otimes H^*(b, k)$ la représentation ρ_1, comme on doit le faire dans la condition 1° du Lemme 5) on obtiendra donc un espace isomorphe à celui qu'on obtiendrait en prenant la représentation ρ_2. Pour cette dernière, on a un isomorphisme de $H^j(g/b, V(g/b) \otimes H^*(b, k))$ sur $H^j(g/b, V(g/b)) \otimes H^*(b, k)$ pour tout entier $j \geqq 0$. Or g/b est abélienne, donc $H^j(g/b, V(g/b)) = (0)$ pour tout $j > 0$ (Théorème 3). La condition 1° du Lemme 5 est donc également vérifiée, ce qui prouve l'assertion sur σ^*_v. Lorsque de plus g est nilpotente, la représentation de g/b dans $H^*(b, k)$ est une représentation par endomorphismes nilpotents et les Lemmes 4 et 5 permettent de démontrer de la même manière l'assertion sur σ^*_w.

5. Filtrations des algèbres enveloppantes. Soit g une algèbre de Lie sur k et soit b une *sous-algèbre* de g. L'isomorphisme $b \to g$ se prolonge en un isomorphisme de $E(b)$ sur la sous-algèbre de $E(g)$ engendrée par l'unité I et les éléments de $b \subset g$. On va dans les paragraphes suivants *identifier* $E(b)$ avec cette sous-algèbre de $E(g)$. L'inclusion $E(b) \subset E(g)$ définit dans $E(g)$ une structure de module à gauche sur $E(b)$. Il résulte du Théorème de Poincaré-Birkhoff [14] que si $x_1, x_2, \cdots x_n$ est une base d'un supplémentaire de b dans g, les monômes de la forme $x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n}$ où les p_i sont des entiers $\geqq 0$, constituent une *base* de $E(g)$ sur $E(b)$.

On suppose désormais que b est un *idéal* de g. Soit A l'idéal bilatère de $E(b)$ engendré par b et soit N le noyau de l'homomorphisme canonique de $E(g)$ sur $E(g/b)$. On peut encore définir N comme l'idéal à gauche (ou

[14] Cf. G. Birkhoff, "Representability of Lie algebras and Lie groups by matrices," *Annals of Mathematics*, 38 (1937), pp. 526-532.

à droite) engendré par b dans $E(g)$. On a donc $N = E(g)A = AE(g)$ et, plus généralement, $N^p = E(g)A^p = A^pE(g)$ pour tout entier $p > 0$. On posera $A^0 = E(b)$ et $N^0 = E(g)$.

LEMME 6. *Etant donnée une base de $E(g)$ considéré comme module à gauche sur $E(b)$, pour qu'un élément de $E(g)$ soit dans N^p, il faut et il suffit qu'il soit combinaison linéaire à coefficients dans A^p des éléments de base.*

La condition est suffisante car $A^p \subset N^p$. Elle est nécessaire car, si $c \in N^p = A^pE(g)$, on peut écrire $c = \sum_i a_i e_i$ avec $a_i \in A^p$ et $e_i \in E(g)$. En exprimant chaque e_i comme combinaison linéaire à coefficients dans $E(b)$ des éléments de base, on obtient pour c une combinaison linéaire à coefficients dans A^p de ces éléments de base.

A la famille d'idéaux N^p correspond une famille de sous-modules C_p dans le g-module $\mathfrak{L}(E(g), k)$, chaque C_p étant le sous-module des fonctions nulles sur N^p. On a $(0) = C_0 \subset C_1 \cdots C_p \subset \cdots$, mais $\mathfrak{L}(E(g), k)$ *ne sera pas la réunion des* C_p. On étudiera dans ce qui suit les modules quotients C_p/C_{p-1}.

A toute fonction $f \in C_p$ $(p > 0)$, on associe l'application linéaire $F: A^{p-1} \to \mathfrak{L}(E(g), k)$ définie par $(F \cdot c) \cdot e = f \cdot (ce)$ quels que soient $c \in A^{p-1}$ et $e \in E(g)$. Si l'on définit dans l'espace $\mathfrak{L}(A^{p-1}, \mathfrak{L}(E(g), k))$ des applications k-linéaires de A^{p-1} dans $\mathfrak{L}(E(g), k)$ une structure de g-module en posant $(xF) \cdot a = x(F \cdot a)$ $(x \in g, a \in A^{p-1})$, l'application $f \to F$ devient un homomorphisme de g-module. Puisque $f \cdot N^p = (0)$, F est nulle sur A^p et peut donc être considérée comme élément du sous-module $\mathfrak{L}(A^{p-1}/A^p, \mathfrak{L}(E(g), k))$. D'autre part, si $e \in N$, alors $(F \cdot c) \cdot e = 0$ pour tout $c \in A^{p-1}$. Par conséquent, $F \cdot c$ est dans $\mathfrak{L}(E(g/b), k)$ considéré comme sous-module de $\mathfrak{L}(E(g), k)$. On a donc obtenu ainsi pour tout entier $p > 0$ un homomorphisme de g-modules

$$\Gamma_p : C_p \to \mathfrak{L}(A^{p-1}/A^p, \mathfrak{L}(E(g/b), k)).$$

Si $f \in C_{p-1}$, alors $\Gamma_p \cdot f = 0$. Réciproquement, si $f \in C_p$ et si $\Gamma_p \cdot f = 0$, alors, pour tout $c \in N^{p-1}$, donc de la forme $\sum_i a_i e_i$ avec $a_i \in A^{p-1}$ et $e_i \in E(g)$, on a $f \cdot c = \sum_i f \cdot (a_i e_i) = \sum_i ((\Gamma_p \cdot f) \cdot a'_i) \cdot e'_i = 0$, où a'_i et e'_i désignent respectivement les images canoniques de a_i et e_i dans les quotients A^{p-1}/A^p et $E(g/b)$. Par suite $f \in C_{p-1}$. Pour tout entier $p > 0$, C_{p-1} est donc le noyau de Γ_p. On va maintenant démontrer que pour tout $F \in \mathfrak{L}(A^{p-1}/A^p, \mathfrak{L}(E(g/b), k))$ il existe un $f \in C_p$ tel que $\Gamma_p \cdot f = F$. Soit en effet (x_i) $(i = 1, 2 \cdots n)$ une base d'un supplémentaire de b dans g. Tout élément $e \in N^{p-1}$ s'écrit d'une manière et d'une seule $e = \sum_j a_j m_j$ où les $a_j \in A^{p-1}$ et où les m_j sont des

monômes de la forme $x_1^{p_1}x_2^{p_2}\cdots x_n^{p_n}$ (Lemme 6). On peut donc définir une fonction $f'\colon N \to k$ en posant $f'\cdot e = \sum_j (F\cdot a'_j)\cdot m'_j$ où a'_j désigne l'image canonique de a_j dans A^{p-1}/A^p et m'_j l'image canonique de m_j dans $E(g/b)$. La fonction f' est évidemment nulle sur N^p. Soit f une fonction dans $\mathfrak{L}(E(g), k)$ ayant f' pour restriction à N^{p-1}; c'est un élément de C_p dont l'image par Γ_p est F. On notera désormais γ_p *l'isomorphisme de* C_p/C_{p-1} *sur* $\mathfrak{L}(A^{p-1}/A^p, \mathfrak{L}(E(g/b), k))$, déduit de Γ_p par passage au quotient.

7. Filtration des modules $V(g)$ et $W(g)$. Les notations étant celles du paragraphe précédent, on posera $V_p(g) = C_p \cap V(g)$ et $W_p(g) = C_p \cap W(g)$ pour tout entier $p > 0$.

LEMME 7. *Le module $W(g)$ est réunion des modules $W_p(g)$. Si g est une algèbre de Lie résoluble sur un corps de caractéristique 0 et si b en est l'idéal dérivé, alors $V(g)$ est réunion des $V_p(g)$.*

En effet, si $f \in W(g)$, c'est une fonction linéaire nulle sur un idéal bilatère de la forme $A(g)^r$ dans $E(g)$, donc à fortiori nulle sur N^r puisque $N \subset A(g)$. D'autre part, on sait que si g est une algèbre de Lie résoluble sur un corps de caractéristique 0, pour tout g-module M de dimension finie sur ce corps, les endomorphismes de l'espace M qui correspondent à des éléments de l'idéal dérivé de g sont dans le radical de l'algèbre enveloppante de M.[15] Ceci signifie que, si D est un idéal bilatère de codimension finie dans $E(g)$, l'image inverse dans $E(g)$ du radical de l'algèbre associative $E(g)/D$ contient b, donc également N. Il en résulte que tout idéal bilatère de codimension finie dans $E(g)$ contient un idéal de la forme N^r ce qui prouve que tout élément de $V(g)$ est dans un sous-module $V_r(g)$.

La structure de g-module de $\mathfrak{L}(E(g/b), k)$ étant définie par sa structure canonique de g/b-module et l'homomorphisme de $E(g)$ sur $E(g/b)$, les éléments de caractère fini (respectivement nilpotent) du g-module $\mathfrak{L}(E(g/b), k)$ coïncident avec les éléments de caractère fini (respectivement nilpotent) de $\mathfrak{L}(E(g/b), k)$ muni de sa structure canonique de g/b-module, c'est à dire les éléments de $V(g/b)$ (respectivement $W(g/b)$). Les quotients A^{p-1}/A^p étant de dimension finie sur k, les éléments de caractère fini (respectivement nilpotent) du g-module $\mathfrak{L}(A^{p-1}/A^p, \mathfrak{L}(E(g/b), k))$ sont donc les éléments du sous-module $\mathfrak{L}(A^{p-1}/A^p, V(g/b))$ (respectivement $\mathfrak{L}(A^{p-1}/A^p, W(g/b))$). D'un autre côté, $V_p(g)/V_{p-1}(g)$ s'identifie à un sous-module du module des éléments de carac-

[15] Cf. C. Chevalley, "Algebraic Lie algebras," *Annals of Mathematics*, 48 (1947), pp. 91-100.

tère fini de C_p/C_{p-1}. De même, $W_p(g)/W_{p-1}(g)$ s'identifie à un sous-module du module des éléments de caractère nilpotent de C_p/C_{p-1}. L'isomorphisme γ_p du paragraphe précédent applique donc *isomorphiquement* $V_p(g)/V_{p-1}(g)$ *dans* $\mathfrak{L}(A^{p-1}/A^p, V(g/b))$ et $W_p(g)/W_{p-1}(g)$ *dans* $\mathfrak{L}(A^{p-1}/A^p, W(g/b))$.

LEMME 8. *Si b est un idéal de codimension 1 dans une algèbre nilpotente g, alors, pour tout $p>0$, γ_p applique isomorphiquement $W_p(g)/W_{p-1}(g)$ sur $\mathfrak{L}(A^{p-1}/A^p, W(g/b))$.*

Le seul point restant à démontrer est que tout élément de caractère nilpotent dans le g-module C_p/C_{p-1} est dans $W_p(g)/W_{p-1}(g)$. Soit f un élément de C_p ayant pour image dans C_p/C_{p-1} un élément de caractère nilpotent. Ceci signifie que f est une fonction dans $\mathfrak{L}(E(g), k)$ nulle sur N^p et sur un idéal de la forme $N^{p-1}A(g)^m$ où m désigne un entier $\geqq 0$. On va démontrer que, quels que soient les entiers p et $m \geqq 0$, il existe un entier $r \geqq 0$ tel que $N^{p-1} \cap A(g)^r \subset N^p + N^{p-1}A(g)^m$. Il en résultera bien que la restriction de f à N^{p-1} est aussi la restriction d'une fonction $f' \in \mathfrak{L}(E(g), k)$ nulle sur $A(g)^r$, donc que $f \in W_p(g) + C_{p-1}$. Soit x un élément de $g - b$. Quel que soit l'entier $r > 0$, tout élément $e \in A(g)^r$ s'écrit, modulo N^p, comme somme de monômes ayant au moins $r - p + 1$ facteurs x, les autres facteurs étant dans b et en nombre au plus égal à $p - 1$. Puisque les x^q constituent une base de $E(g)$ considéré comme module à gauche sur $E(b)$ (cf. paragraphe 5), on pourra écrire chacun de ces monômes sous la forme $\sum_q c_q x^q$ avec $c_q \in E(b)$. Pour cela, on devra effectuer un certain nombre de fois l'opération consistant à permuter un facteur x avec un facteur $y \in b$, ce qui fait apparaître chaque fois un monôme supplémentaire ayant un facteur x en moins, mais dont l'un des facteurs $y \in b$ aura été remplacé par $[x, y] = ad(x) \cdot f$. Puisque g est supposée nilpotente, il existe un entier s tel que $ad(x)^s = 0$. Les monômes dont on part comportant au plus $p - 1$ facteurs dans b, on voit que le nombre des facteurs x restera au moins égal à $r - p + 1 - s(p-1)$. Ainsi, tout élément $e \in A(g)^r$ est égal, modulo N^p à une somme de termes de la forme $c_q x^q$ avec $q \geqq r-(s-1)(p-1)$ et $c_q \in E(b)$. Si de plus $e \in N^{p-1}$, alors les c_q sont dans A^{p-1} (Lemme 6). Par conséquent, en choisissant $r \geqq m + (s-1)(p-1)$, on a $N^{p-1} \cap A(g)^r \subset N_p + N^{p-1}A(g)^m$, ce qui achève la démonstration.

LEMME 9. *Si b est l'idéal dérivé d'une algèbre de Lie g, alors γ_p applique isomorphiquement $V_p(g)/V_{p-1}(g)$ sur $\mathfrak{L}(A^{p-1}/A^p, V(g/b))$.*

Soit (x_i) $(i = 1, 2, \cdots n)$ une base d'un supplémentaire de b dans g. On notera $S(i, i+r)$ $(i = 1, 2, \cdots n;\ r = 0, 1, \cdots n - i)$ le sous-espace de $E(g)$ engendré par les monômes de la forme $x_i^{p_i} x_{i+1}^{p_{i+1}} \cdots x_{i+r}^{p_{i+r}}$ où les

$p_i, p_{i+1} \cdots p_{i+r}$ sont des entiers $\geqq 0$. On posera pour abréger $S = S(1, n)$ et $S(i) = S(i, i)$. L'homomorphisme canonique ζ de $E(g)$ sur $E(g/b)$ applique isomorphiquement S sur $E(g/b)$. On utilisera dans la démonstration du Lemme 9 le résultat préliminaire suivant:

LEMME 10. *Pour tout sous-espace $U \subset S$ dont l'image par ζ est un idéal de codimension finie dans $E(g/b)$, il existe un sous-espace $U' \subset U$ tel que*

1° *$\zeta \cdot U'$ est un idéal de codimension finie dans $E(g/b)$.*

2° $U'g \subset U' + bU$.

On remarquera pour commencer que, si $u \in S(i, i+r)$, alors quel que soit $z \in g$,

(A) $zu - uz \in bS(i, i+r)$.

Pour vérifier cette inclusion, on supposera que u est un monôme et l'on procèdera par induction sur le nombre de ses facteurs en posant $u = u'x_{i+s}$ avec $0 \leqq s \leqq r$ et $u' \in S(i, i+s)$. On a

$$zu - uz = (zu' - u'z)\, x_{i+s} + u'(zx_{i+s} - x_{i+s}z) \in bS(i, i+s) + u'b \subset bS(i, i+r)$$

d'après l'hypothèse inductive.

D'autre part, si $u \in S(i, i+r)$ et si $i \leqq j \leqq i+r$, alors

(B) $ux_j \in bS(i, i+r) + S(i, i+r)$.

On supposera de même que u est un monôme et l'on procèdera par induction sur le nombre de ses facteurs en posant $u = u'x_{i+s}$ avec $0 \leqq s \leqq r$ et $u' \in S(i, i+s)$. On a trivialement $ux_j \in S(i, i+r)$ si $j \geqq i+s$. Si $j < i+s$, on écrira

$$ux_j = u'(x_{i+s}x_j - x_jx_{i+s}) + u'x_jx_{i+s} \in u'b + (bS(i, i+s) + S(i, i+s))x_{i+s}$$

d'après l'hypothèse inductive. Ce sous-espace est contenu dans

$$S(i, i+s) + bS(i, i+s) \subset S(i, i+r) + bS(i, i+r)$$

d'après (A).

Ceci étant, soit β l'application linéaire du produit tensoriel (sur k) $g \otimes S(i)$ dans le sous-espace $bS(i) \subset E(g)$ définie par $\beta \cdot (z \otimes u) = zu - uz$ pour $z \in g$ et $u \in S(i)$. Posons $K_i = U \cap S(i)$; c'est un sous-espace de codimension finie dans $S(i)$. Puisque b est de dimension finie, bK_i est de codimension finie dans $bS(i)$, et il existe donc un sous-espace de codimension finie dans $g \otimes S(i)$ dont l'image par β est dans bK_i. Puisque g est de dimen-

sion finie, il existe donc un sous-espace K'_i de codimension finie dans $S(i)$ tel que $\beta\cdot(g\otimes K'_i)\subset bK_i$. Les deux sous-espaces K_i et K'_i de $S(i)$ étant de codimension finie, leur intersection contient un polynôme $P_i\neq 0$ en x_i. Soit U'_i le sous-espace $S(1,i-1)P_iS(i,n)$ de S.[16] Son image par ζ est l'idéal de $E(g/b)$ engendré par $\zeta\cdot P_i$. Puisque $P_i\,\varepsilon\, U$, on a donc $\zeta\cdot U'_i\subset\zeta\cdot U$ d'où $U'_i\subset U$. D'autre part, $U'_ig\subset U'_i+bU$. En effet, si $z\,\varepsilon\, b$, alors $U'_iz\,\varepsilon\, S(1,i-1)P_ibS(i,n)$ d'après (A). Or, pour tout $y\,\varepsilon\, b$, on a $yP_i-P_iy\,\varepsilon\, bK_i\subset bU$, donc $P_ib\subset bU$ et par suite, $\zeta\cdot U$ étant un idéal, $S(1,i-1)P_ibS(i,n)\subset bU$ d'après (A). D'autre part, si $i\leqq j\leqq n$, on a

$$\begin{aligned}U'_ix_j&=S(1,i-1)P_iS(i,n)x_j\\&\subset S(1,i-1)P_iS(i,n)+S(1,i-1)P_ibS(i,n)\end{aligned}$$

d'après (B), donc $U'_ix_j\subset U'_i+bU$. Enfin, si $j<i$,

$$U'_ix_j\subset S(1,i-1)P_ix_jS(i,n)+S(1,i-1)P_ibS(i,n)$$

d'après (A). Puisque $P_ix_j-x_jP_i\,\varepsilon\, bK_i$ on a donc

$$U'_ix_j\subset S(1,i-1)x_jP_iS(i,n)+bU$$

d'après (A) et ce sous-espace est contenu dans U'_i+bU d'après (B). Soit alors $U'=\sum_i U'_i$; son image $\zeta\cdot U'$ dans $E(g/b)$ est l'idéal engendré par les $\zeta\cdot P_i$. Puisque $E(g/b)$ est l'algèbre commutative des polynômes en $\zeta\cdot x_i$, l'idéal $\zeta\cdot U'$ est de codimension finie dans $E(g/b)$. D'autre part, $U'\subset U$ et $U'g\subset U'+bU$ par suite des inclusions analogues vérifiées par chaque U'_i. Ceci termine la démonstration du Lemme 10.

Revenons maintenant à la *démonstration du Lemme* 9. Il reste à vérifier que tout élément de caractère fini dans C_p/C_{p-1} est l'image d'un élément de caractère fini de C_p. Soit f un élément de C_p ayant pour image dans C_p/C_{p-1} un élément de caractère fini. Ceci signifie que f est une fonction dans $\mathfrak{L}(E(g),k)$ nulle sur N^p et sur un idéal de la forme $N^{p-1}H$ où H est un idéal bilatère de codimension finie dans $E(g)$. On va démontrer que, quel que soit l'entier p et l'idéal H de codimension finie dans $E(g)$, *il existe un idéal bilatère H' de codimension finie dans $E(g)$ tel que* $N^{p-1}\cap H'\subset N^p+N^{p-1}H$. Il en résultera bien que la restriction de f à N^{p-1} est aussi la restriction à N^{p-1} d'une fonction $f'\,\varepsilon\,\mathfrak{L}(E(g),k)$ nulle sur H, donc contenue dans $V_p(g)$. Soit U_1 l'intersection de S et de H. Le Lemme 10 permet d'y associer une suite de sous-espaces U_k $(k=0,1,\cdots)$ possédant les propriétés suivantes

1) $S=U_0\supset U_1\supset U_2\cdots U_k\supset U_{k+1}\cdots$,

[16] Pour $i=1$, $S(i,i-1)$ représentera le corps des scalaires.

2) quel que soit k, l'image canonique de U_k dans $E(g/b)$ est un idéal de codimension finie,

3) quel que soit $k \geqq 1$, $U_k g \subset U_k + bU_{k-1}$.

Ceci étant, soit $L_k = A^{p-k}U_k$ $(k = 0, 1, 2, \cdots p)$. Le sous-espace $L = \sum_{k=0}^{k=p} L_k$ est un *idéal à droite* de $E(g)$. En effet,

$$L_k g = A^{p-k}U_k g \subset L_k + A^{p-k}bU_{k-1} = L_k + A^{p-k+1}U_{k-1} \subset L_k + L_{k-1}$$

pour $k \geqq 1$, et pour $k = 0$, on a $L_0 = A^p U_0 = A^p S = N^p$. D'autre part, $L_0 + L_p$ est de codimension finie dans $E(g)$. En effet, si (t_j) est une base d'un supplémentaire de U_p dans S, tout élément de $E(g)$ s'écrira modulo $L_p = E(b)U_p$ comme combinaison linéaire à coefficients dans $E(b)$ des t_j. Puisque pour $a_j \in A^p$, $a_j t_j \in N^p = L_0$ et que A^p est de codimension finie dans $E(b)$, l'espace $L_0 + L_p$ est bien de codimension finie dans $E(g)$. A fortiori, l'idéal L *est de codimension finie dans* $E(g)$. On va enfin montrer que l'intersection $L \cap N^{p-1}$ est $L_0 + L_1 = N^p + A^{p-1}U \subset N^p + N^{p-1}H$. Tout élément de L_k $(k = 0, 1, \cdots p)$ s'écrit comme combinaison linéaire à coefficients dans A^{p-k} d'éléments linéairement indépendants de U_k, et s'il est dans N^{p-k+1} ces coefficients sont dans A^{p-k+1} d'après le Lemme 6. On a donc $N^{p-k+1} \cap L_k \subset A^{p-k+1}U_k \subset L_{k-1}$ pour tout $k > 0$. D'autre part, $L_k \subset N^{p-k}$ pour $0 \leqq k \leqq p$. Il en résulte que $L_p \cap N \subset L_{p-1}$ et $(L_p + L_{p-1}) \cap N \subset L_{p-1}$ donc $(L_p + L_{p-1}) \cap N^2 \subset L_{p-1} \cap N^2 \subset L_{p-2}$ et par suite $(L_p + L_{p-1} + L_{p-2}) \cap N^2 \subset L_{p-2}$. De proche en proche, on démontrera ainsi que $(L_p + L_{p-1} \cdots L_1) \cap N^{p-1} \subset L_1$, donc que $L \cap N^{p-1} \subset L_1 + L_0$. On a donc obtenu un idéal à droite L de codimension finie dans $E(g)$ dont l'intersection avec N^{p-1} est dans $N^p + N^{p-1}H$. Puisque tout idéal à droite de codimension finie contient un idéal bilatère de codimension finie, le Lemme 9 est ainsi démontré.

7. Trivialité de $H^*(g, W(g))$ lorsque g est nilpotente.

THÉORÈME 6. *Si g est une algèbre de Lie nilpotente, alors $H^p(g, W(g)) = (0)$ pour tout $p > 0$.*

La démonstration procède par induction sur la dimension de g. Le Théorème est vérifié pour g de dimension 1 (Théorème 4). Supposons le vérifié pour les algèbres nilpotentes de dimension inférieure à la dimension de l'algèbre nilpotente g. Soit b un idéal de codimension 1 dans g et soit η l'homomorphisme de $W(g)$, considéré comme b-module, dans $W(b)$ qui associe à tout $f \in W(g) \subset \mathfrak{L}(E(g), k)$ sa restriction à $E(b) \subset E(g)$. L'homomorphisme η se prolonge en un homomorphisme $\bar{\eta}$ du complexe de cochaînes $C(b, W(g))$ dans $C(b, W(b))$. En composant l'homomorphisme canonique $C(g, W(g)) \to C(b, W(g))$ avec $\bar{\eta}$, on obtient un homomorphisme

$\xi: C(g, W(g)) \to C(b, W(b))$. Compte tenu de l'hypothèse inductive, il suffit de démontrer que cet homomorphisme induit un isomorphisme de $H^*(g, W(g))$ sur $H^*(b, W(b))$. On utilisera pour cela des filtrations de ces deux complexes. Dans $C(g, W(g))$, on prendra la filtration par les sous-complexes $C(g, W_p(g))$ correspondant à la filtration du module $W(g)$ par les $W_p(g)$ relatifs à b (Lemme 7). Dans $C(b, W(b))$, on utilisera la filtration par les sous-complexes $C(b, W_p(b))$ où chaque $W_p(b)$ est le sous-module des fonctions linéaires sur $E(b)$ nulles sur l'idéal A^p. Pour tout $p \geqq 0$, on a $\eta \cdot W_p(g) \subset W_p(b)$ et par suite $\xi \cdot C(g, W_p(g)) \subset C(b, W_p(b))$. D'après le Lemme 8, l'isomorphisme γ_p permet d'identifier $C(g, W_p(g))/C(g, W_{p-1}(g)) = C(g, W_p(g)/W_{p-1}(g))$ avec $C(g, \mathfrak{L}(A^{p-1}/A^p, W(g/b))$ qui est aussi le complexe $\mathfrak{L}(A^{p-1}/A^p, C(g, W(g/b))$ des applications k-linéaires de l'espace A^{p-1}/A^p dans $C(g, W(g/b))$. D'autre part, $C(b, W_p(b))/C(b, W_{p-1}(b))$ s'identifie au complexe $\mathfrak{L}(A^{p-1}/A^p, C(b, k))$. Après ces identifications, l'homomorphisme déduit de ξ par passage au quotient devient l'application de $\mathfrak{L}(A^{p-1}/A^p, C(g, V(g/b))$ dans $\mathfrak{L}(A^{p-1}/A^p, C(b, k))$ qui est définie par l'homomorphisme $\sigma_w: C(g, W(g/b)) \to C(b, k)$ qui a été étudié au paragraphe 4. Compte tenu du Théorème 5, cet homomorphisme induit un isomorphisme de la cohomologie du premier complexe sur la cohomologie du second. Il résulte donc du " Lemme des cinq " que ξ induit bien un isomorphisme de $H^*(g, W(g))$ sur $H^*(b, W(b))$.

8. Trivialité de $H^*(g, V(g))$ lorsque g est résoluble.

THÉORÈME 7. *Si g est une algèbre de Lie résoluble sur un corps de caractéristique 0, alors $H^p(g, V(g)) = (0)$ pour tout entier $p > 0$.*

La démonstration consiste à se ramener au Théorème 6. Soit b l'idéal dérivé de g. C'est une algèbre de Lie nilpotente. Tout élément de $V(g)$ étant contenu dans un sous-module $V_p(g)$ relatif à l'idéal b (Lemme 7), sa restriction à $E(b)$ sera nulle sur un idéal $A^p \subset N^p$ et c'est donc un élément de $W(b)$. On a ainsi un homomorphisme canonique η de $V(g)$ considéré comme b-module dans $W(b)$ et cet homomorphisme se prolonge en un homomorphisme $\overline{\eta}: C(b, V(g)) \to C(b, W(b))$. En composant l'homomorphisme canonique $C(g, V(g)) \to C(b, V(g))$ avec $\overline{\eta}$, on obtient un homomorphisme canonique du complexe $C(g, V(g))$ dans le complexe $C(b, W(b))$. Compte tenu du Théorème 6, il suffit de démontrer que cet homomorphisme induit un isomorphisme de $H^*(g, V(g))$ sur $H^*(b, W(b))$. On utilisera pour cela des filtrations de ces deux complexes. Dans $C(g, V(g))$ on prendra la filtration par les sous-complexes $C(g, V_p(g))$ (Lemme 7) et dans $C(b, W(b))$ la filtration par les sous-complexes $C(b, W_p(b))$ déja introduits dans la démonstration du

Théorème 6. Pour tout $p \geqq 0$, on a $\eta \cdot V_p(g) \subset W_p(b)$ et par suite $\xi \cdot C(g, V_p(g)) \subset C(b, W_p(b))$.

D'après le Lemme 9, l'isomorphisme γ_p permet d'identifier

$$C(g, V_p(g))/C(g, V_{p-1}(g)) = C(g, V_p(g)/V_{p-1}(g))$$

avec $C(g, \mathfrak{L}(A^{p-1}/A^p, V(g/b))$ qui est aussi le complexe $\mathfrak{L}(A^{p-1}/A^p, C(g, V(g/b))$ des applications k-linéaires de l'espace A^{p-1}/A^p dans $C(g, V(g/b))$. D'autre part $C(b, W_p(b))/C(b, W_{p-1}(b))$ s'identifie au complexe $\mathfrak{L}(A^{p-1}/A^p, C(b, k))$. Après ces identifications, l'homomorphisme déduit de ξ par passage au quotient devient l'homomorphisme de $\mathfrak{L}(A^{p-1}/A^p, C(g, V(g/b))$ dans $\mathfrak{L}(A^{p-1}/A^p, C(b, k))$ qui est défini par l'homomorphisme σ_v: $C(g, V(g/b)) \to C(b, k)$ étudié au paragraphe 4. Compte tenu du Théorème 5, cet homomorphisme induit un isomorphisme de la cohomologie du premier complexe sur la cohomologie du second. Il résulte donc du "Lemme des cinq" que ξ induit bien un isomorphisme de $H^*(g, V(g))$ sur $H^*(b, W(b))$.

Démonstration du Théorème I. Soit M un module de représentation d'une algèbre de Lie résoluble g sur un corps de caractéristique 0. Pour tout entier $p > 0$, on a $H^p(g, \mathfrak{L}_f(E(g), M)) = H^p(g, V(g)) \otimes M = (0)$ d'après le Théorème 7. Compte tenu du Théorème 2, ceci prouve que les classes dans $H^p(g, M)$ sont effaçables.

On observera pour terminer que, compte tenu du Théorème 3, le Théorème 8 signifie que, sur un groupe résoluble simplement connexe, donc homéomorphe à R^n, la cohomologie du complexe des formes différentielles réelles dont les translatées sont dans un espace de dimension finie est la même que celle du complexe de toutes les formes différentielles. Nous ne connaissons pas de démonstration directe de ce résultat. Le même circonstance se vérifie trivialement pour un groupe compact; elle n'est cependant pas générale.

UNIVERSITÉ DE STRASBOURG.

REFERENCES.

[1] C. Chevalley et S. Eilenberg, "Cohomology theory of Lie groups and Lie algebras," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 85-124.

[2] G. Hochschild et J. P. Serre, "Cohomology of Lie algebras," *Annals of Mathematics*, vol. 57 (1953), pp. 591-603.

[3] Iwasawa, "On the representations of Lie algebras," *Japanese Journal of Mathematics*, vol. 19 (1948), pp. 405-426.

COHOMOLOGY OF RESTRICTED LIE ALGEBRAS.*

By G. HOCHSCHILD.

Introduction. We shall be concerned with restricted Lie algebras of characteristic p ($\neq 0$), in the sense of N. Jacobson, [6]. The structure of a restricted Lie algebra L comprises, in addition to the usual Lie algebra structure, a map of L into itself, denoted $x \to x^{[p]}$, with properties that correspond, in a fashion we shall make precise later, to the properties of the map $x \to x^p$ in an associative algebra of characteristic p. The appropriate representation theory for such algebras is accordingly obtained by confining one's attention to the *restricted* representations in which the transformation corresponding to $x^{[p]}$ is the p-th power of the transformation corresponding to x. Just as the ordinary representations of a Lie algebra correspond to the representations of its universal enveloping (or Birkhoff-Witt) algebra, the restricted representations of a restricted Lie algebra correspond to the representations of a certain homomorphic image of its universal enveloping algebra; its u-algebra, in the sense of Jacobson, [6]. It has been shown by H. Cartan and S. Eilenberg that the cohomology groups of a Lie algebra can be computed, as described in section 1 below, from any augmented free acyclic complex over its universal enveloping algebra. If, in this context, one replaces the universal enveloping algebra by the u-algebra one obtains new groups, which we shall call the restricted cohomology groups. One is thus led to the problem of determining what becomes of the usual elementary interpretations of the 1- and 2-dimensional cohomology groups in terms of extensions of modules and Lie algebras when one passes from the "ordinary" situations to the corresponding "restricted" ones. This is what we propose to do here.[1] The results are what one would naturally expect them to be, except perhaps for those which deal with the relations between the "ordinary" objects and the corresponding "restricted" ones, such as are expressed in Theorems 3. 1 and 3. 2.

---

* Received October 14, 1953. Part of this work was done while the author was participating in the conference on Lie algebras and Lie groups, July-August, 1953, organized by the American Mathematical Society with the financial support of the National Science Foundation.

[1] I am indebted to N. Jacobson for the suggestion that restricted Lie algebras can thus be used as experimental animals to be observed in the environment of general cohomology theory.

1. Basic notions. Let L be a Lie algebra over a field F. We denote by T_L the tensor algebra built over the F-space L.[2] As an F-space, T_L is the direct sum of subspaces $T_L{}^i$, $i = 0, 1, \cdots$, where $T_L{}^0 = F$, $T_L{}^1 = L$, $T_L{}^2 = L \otimes L$, and generally $T_L{}^i$ is the i-fold tensor product $L \otimes \cdots \otimes L$, relative to F, of L by itself. The multiplication in T_L is defined by the canonical map $(T_L{}^i, T_L{}^j) \to T_L{}^i \otimes T_L{}^j = T_L{}^i T_L{}^j = T_L{}^{i+j}$. Let I_L denote the 2-sided ideal of T_L which is generated by the elements of the form $x \otimes y - y \otimes x - [x, y] \in L \otimes L + L$, with x and y ranging over L, and $[x, y]$ denoting their commutator in L. The algebra $R_L = T_L/I_L$ is the universal enveloping algebra of L. We identify F with its natural isomorphic image $(F + I_L)I_L$ in R_L, and similarly we identify L with its natural isomorphic image $(L + I_L)/I_L$ in R_L.

The natural image $R_L{}^+$ of $\sum_{i=1}^{\infty} T_L{}^i$ in R_L is a 2-sided ideal in R_L and, as a vector space over F, R_L is the direct sum of $R_L{}^+$ and F. We have a canonical projection ϵ of R_L onto F whose kernel is $R_L{}^+$, namely the map induced by the projection of T_L onto $T_L{}^0$.

The importance of R_L lies in the almost evident fact that the representations of L correspond in a natural 1-1 fashion to the "unitary" representations of R_L, i. e., to those in which the identity element of R_L is mapped into the identity transformation. Accordingly, the cohomology theory of L can be based entirely on the ring R_L, with the projection ϵ, as has in fact been done by H. Cartan and S. Eilenberg. Since their general theory of algebraic complexes has not yet appeared in print, we shall be content with only a few allusions to it, which may eventually serve to clarify the underlying ideas for some of our manipulations of the cohomology groups.

We regard F as an R_L-module with "trivial operators," i. e., if $r \in R_L$ and $\alpha \in F$, we set $r \cdot \alpha = \epsilon(r)\alpha$. The usual cohomology groups $H^n(L, M)$ for L in an L-module M coincide with the groups $\mathrm{Ext}_{R_L}{}^n(F, M)$, as defined by Cartan-Eilenberg, where M is regarded as a unitary R_L-module in the natural fashion.

The usual explicit computation of $H(L, M)$ starts with the alternating n-linear functions on L with values in M as the n-cochains, and with the coboundary operator δ, where

$$(\delta f)(x_0, \cdots, x_n) = \sum_{i=0}^{n} (-1)^i x_i \cdot f(x_0, \cdots, x_n) + \sum_{p<q} (-1)^{p+q} f([x_p, x_q], x_0, \cdots, x_n).^3$$

[2] We look upon this and related concepts from the point of view of chapter 1 of [1]. In particular, see p. 4 of [1] for a precise definition of the tensor algebra.

[3] For the details of this description of $H(L, M)$, see § 1 of [5].

It is a consequence of the general theory of algebraic complexes that the following cochains and coboundary operator lead to the same cohomology groups, up to a canonical isomorphism. The n-cochains are now the n-linear functions (not alternating) on R_L^+ with values in M, and the coboundary operator δ is defined by the formula

$$(\delta g)(s_0, \cdots, s_n) = s_0 \cdot g(s_1, \cdots, s_n) + \sum_{i=1}^{n} (-1)^i g(s_0, \cdots, s_{i-1}s_i, \cdots, s_n).$$

Observe that this is what one obtains from the cohomology theory of associative algebras if one makes all the right operators 0.

One can give an explicit cochain map which induces the canonical isomorphism between the cohomology groups as defined above in the two different ways. For every cochain g of the second, "associative" type, define a cochain g' of the first, "Lie" type by the formula

$$g'(x_1, \cdots, x_n) = \sum_{\sigma} e_\sigma g(x_{\sigma(1)}, \cdots, x_{\sigma(n)}),$$

where the summation is over all permutations σ of the set $(1, \cdots, n)$, and where e_σ is equal to $+1$ or -1 according to whether σ is even or odd. It can be verified directly that $\delta(g') = (\delta g)'$, so that the map $g \to g'$ induces a homomorphism of the cohomology groups. The fact that this homomorphism is actually an isomorphism onto is proved by means of the general theory of algebraic complexes. In order to apply this theory, we must observe that, in the terminology of Eckmann, [2], each of the above descriptions of $H(L, M)$ exhibits it as the "equivariant" cohomology group of some (augmented) free acyclic R_L-complex. The first description is obtained from the following complex: let $C_0 = R_L$, regarded as an R_L-module in the natural fashion. This is the 0-dimensional part of the chain group, regarded as an R_L-module. The augmentation is $\epsilon: C_0 \to F$, as defined above. For $n > 0$, denote by L_n the n-fold exterior product of the F-space L by itself.[4] The n-dimensional part of our R_L-module of chains is defined as $C_n = R_L \otimes L_n$, with its natural structure of an R_L-module. The boundary operator d of our complex is characterized by the formula

$$d(r \otimes (x_0 \wedge \cdots \wedge x_n))$$

$$= \sum_{i=0}^{n} (-1)^i (rx_i) \otimes (x_0 \wedge \cdots \wedge x_n) + \sum_{p<q} (-1)^{p+q} r \otimes ([x_p, x_q] \wedge x_0 \wedge \cdots \wedge x_n).$$

Evidently, each C_n is a free R_L-module, and it is not too difficult to verify that

[4] This means that L_n is the homogeneous component of degree n of the exterior algebra over L, in the sense of § 5 of [1]. The symbol $\wedge$ in the boundary formula below indicates multiplication in this exterior algebra.

$d^2 = 0$, so that we have indeed a free R_L-complex. In order to show that this complex is acyclic, one observes that, by means of an ordered basis of L, one can define a graduation of R_L; the degrees being the total degrees in the basis elements of L in the representation of the elements of R_L as linear combinations of ordered monomials (which are well known to be linearly independent) in these basis elements. One then defines an increasing filtration of the complex by means of a sequence of subcomplexes defined by the condition that, in the m-th complex, the n-chains shall have their coefficients in R_L of degree no greater than $m - n$. The factor complex obtained from two successive such subcomplexes can be described quite simply in terms of the symmetric algebra [5] and the exterior algebra over L, and can be shown to be acyclic by a reductive verification. By a familiar argument of algebraic homology theory, it follows then immediately that the whole complex is acyclic. Now if M is any unitary R_L-module, one constructs the cochain complex whose n-dimensional cochains are the R_L-homomorphisms of C_n into M, and whose coboundary operator δ is the dual of d. The corresponding cohomology group is the "equivariant" cohomology group derived from M and our complex (C, d). It is evident that this gives precisely the usual description of $H(L, M)$ in terms of Lie cochains; one merely has to observe that the linear maps of L_n into M may be identified with the n-linear alternating functions on L with values in M.

The situation of the second, associative description of $H(L, M)$ is a good deal simpler. We construct an appropriate augmented free acyclic R_L-complex (D, d) as follows. $D_0 = C_0$, and the augmentation is the same map ϵ which we have used above. For $n > 0$, $C_n = R_L \otimes R_L^+ \otimes \cdots \otimes R_L^+$ (one factor R_L, followed by n factors R_L^+). The boundary operator is characterized by the formula

$$d(r \otimes s_0 \otimes \cdots \otimes s_n)$$
$$= rs_0 \otimes s_1 \otimes \cdots \otimes s_n + \sum_{i=1}^{n} (-1)^i r \otimes s_0 \otimes \cdots \otimes s_{i-1}s_i \otimes \cdots \otimes s_n.$$

Here we can define an explicit homotopy operator h which shows at once that (D, d) is acyclic. On F, h is defined as the injection into R_L, which is an inverse of ϵ. On $D_0 = R_L$, h is the map $r \to 1 \otimes r^+$, where r^+ denotes the component $r - \epsilon(r)$ of r in R_L^+. Generally,

$$h(r \otimes s_1 \otimes \cdots \otimes s_n) = 1 \otimes r^+ \otimes s_1 \otimes \cdots \otimes s_n.$$

It is easily verified that $dh + hd$ is the identity map of D onto itself, whence

[5] Loosely speaking, this is the algebra of the commutative polynomials in the basis elements of L, with coefficients in F. For a precise definition, see § 4 of [1].

it is clear that our complex is acyclic, while it is evidently free. It is also evident that the corresponding equivariant cohomology group in any unitary R_L-module M is exactly what our second, associative description of $H(L, M)$ yields.

Now, quite generally,[6] if C and D are two augmented complexes over an arbitrary operator ring R, both having the same augmentation module, F, and if C is free over R, while D is acyclic, then the identity map of F onto itself can be extended to an admissible map of C into D, i. e., to a dimension preserving R-homomorphism which satisfies the usual commutativity relations with respect to the boundary operators and the augmentation maps. Moreover, any two such admissible maps are equivalent to each other by a homotopy map of C into D. It follows that there is a canonical homomorphism of the cohomology group derived as above from D into the cohomology group derived from C, this homomorphism being induced by the dual of any admissible map of C into D.

In the situation above, it is clear, by interchanging C and D, that the canonical cohomology homomorphism is an isomorphism onto. In order to complete the proof of our above claims concerning the two descriptions of $H(L, M)$, we merely have to observe that our cochain map $g \to g'$ is the dual of the chain map $\gamma: C \to D$ defined as follows. On C_0, γ is the identity map. On C_n, with $n > 0$, γ is the R_L-homomorphism which is characteribed by the formula $\gamma(r \otimes (x_1 \wedge \cdots \wedge x_n)) = \sum_\sigma e_\sigma r \otimes x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(n)}$. It is not difficult to verify that γ is an admissible map of C into D, so that its dual induces the canonical isomorphism of the associatively defined cohomology group onto the usual one.

Before discussing the restricted cohomology groups, we must recall the basic facts concerning restricted Lie algebras. From now on, F will always denote a field of characteristic $p \neq 0$. A restricted Lie algebra over F is a Lie algebra L over F in which there is defined a map $x \to x^{[p]}$ (of L into itself) which is subjected to the following conditions. Let D_x denote the inner derivation of L which is effected by the element $x \in L$, so that $D_x(y) = [x, y]$. Then the conditions are:

(1) $(D_x)^p = D_{x^{[p]}}$;

(2) for $\alpha \in F$ and $x \in L$, $(\alpha x)^{[p]} = \alpha^p x^{[p]}$;

(3) $(x + y)^{[p]} = x^{[p]} + y^{[p]} + \sigma(x, y)$, where $\sigma(x, y)$ is the element of L which, in R_L, coincides with $(x + y)^p - x^p - y^p$.

[6] What follows is a slightly specialized reminder of some of the basic features of the theory of complexes over a ring, as formulated by Eckmann in [2].

It is not evident that $\sigma(x, y)$ actually belongs to L. This is seen by operating in the polynomial ring $R_L[t]$ in one auxiliary variable, t, which commutes with the elements of the coefficient ring R_L. Write

$$(xt + y)^p = x^p t^p + y^p + \sum_{i=1}^{p-1} s_i(x, y) t^i.$$

Then we have $\sigma(x, y) = \sum_{i=1}^{p-1} s_i(x, y)$. On the other hand, differentiating with respect to t, we obtain

$$\sum_{j=0}^{p-1} (xt + y)^j x (xt + y)^{p-1-j} = \sum_{i=1}^{p-1} i s_i(x, y) t^{i-1}.$$

The left hand side is of the form $\sum_{j=0}^{p-1} u^j v u^{p-1-j}$. Let us momentarily replace u and v by independent non-commuting variables and operate in the ring Q of the non-commutative polynomials in u and v with coefficients in F. If D_u denotes the map $w \to uw - wu = D_u(w)$, we have

$$D_u\left(\sum_{j=0}^{p-1} u^j v u^{p-1-j} - D_u{}^{p-1}(v)\right) = 0.$$

In Q, this is easily seen to imply that $\sum_{j=0}^{p-1} u^j v u^{p-1-j} = D_u{}^{p-1}(v)$. This formal result must continue to hold after specialization to $R_L[t]$. Hence we have $\sum_{i=1}^{p-1} i s_i(x, y) t^{i-1} = D_{(xt+y)^{p-1}}(x)$, and a comparison of coefficients shows that $s_i(x, y)$ is a sum of commutators, and so belongs to L. Thus $\sigma(x, y) \in L$. For future reference, it is important to observe that $\sigma(x, y)$ is obtained by formally integrating the polynomial $D_{(xt+y)^{p-1}}(x)$ with respect to t and then putting $t = 1$. Note that this formal integration can be carried out meaningfully, because the polynomial is actually of degree less than $p - 1$.

Now let L be a restricted Lie algebra over F, and denote by P_L the ideal of R_L which is generated by the elements of the form $x^p - x^{[p]}$, with $x \in L$. Note that, by (1), these elements are in the center of R_L. We define $U_L = R_L/P_L$. U_L is what Jacobson has called the u-algebra of L. We can still identify F and L with their isomorphic images in U_L, and we have then $x^p = x^{[p]}$, in U_L, for all $x \in L$.

For restricted Lie algebras L, it is appropriate to concentrate one's attention on the restricted L-modules, i. e., on those L-modules M in which $x^p \cdot m = x^{[p]} \cdot m$, for all $m \in M$ and all $x \in L$.[7] Clearly, the restricted L-

[7] It is implied in our notation that L-modules are simultaneously regarded as R_L-modules. In particular, when we write $x^p \cdot m$, x^p denotes the p-th power of x in R_L.

modules are precisely the unitary U_L-modules. Contrary to R_L, the algebra U_L is of finite dimension over F, whenever L is finite dimensional. In fact, if (x_i) is an ordered basis for L over F, the ordered monomials $x_{i_1}^{e_1} \cdots x_{i_k}^{e_k}$, with $i_1 < \cdots < i_k$ and $0 \leqq e_j < p$, constitute a basis for U_L over F, whether the basis is finite or not.

Our projection ϵ of R_L onto F induces a projection ϵ' of U_L onto F whose kernel is the natural image U_L^+ of R_L^+ in U_L. If F is regarded as a U_L-module with $u \cdot \alpha = \epsilon'(u)\alpha$, then the restricted cohomology groups for L in the restricted L-module M coincide with the groups $\mathrm{Ext}_{U_L}^n(F, M)$ of Cartan-Eilenberg. One can avoid this reference to the general theory by using the explicit cochain complex of the associative type which is obtained from the one we discussed above simply by replacing R_L^+ with U_L^+. Observe that the usual description of cohomology for Lie algebras in terms of Lie cochains leaves no room for such a modification, so that a reformulation was necessary for our purpose.

The canonical homomorphism of R_L onto U_L allows us to regard any U_L-module also as an R_L-module. An acyclic U_L-complex can therefore be regarded also as an acyclic R_L-complex. It follows from our above general remarks concerning algebraic complexes that there is a canonical homomorphism of the restricted cohomology groups $H_*^n(L, M)$ into the ordinary cohomology groups $H^n(L, M)$ which is induced by the dual of any admissible chain map of the free acyclic R_L-complex used for computing $H(L, M)$ into the free acyclic U_L-complex (regarded as an acyclic R_L-complex) used for computing $H_*(L, M)$. An explicit cochain map inducing this homomorphism is given, in the associative description, by $g \to g^0$, where $g^0(s_1, \cdots, s_n) = g(s'_1, \cdots, s'_n)$, with $s_i \in R_L^+$ and s'_i its canonical image in U_L^+.

Let $C(M)$ denote the complex of the cochains for R_L^+ in the restricted L-module M, and let $C^0(M)$ stand for the subcomplex consisting of the cochains of the form g^0, where g is a cochain for U_L^+ in M. Then we have the exact sequence of complexes $(0) \to C^0(M) \to C(M) \to C(M)/C^0(M) \to (0)$, which gives rise to the usual exact sequence connecting their cohomology groups. Since our map $g \to g^0$ is an isomorphism, we may identify $H^n(C^0(M))$ with $H_*^n(L, M)$, whence we have the exact sequence

$$\cdots \to H_*^n(L, M) \to H^n(L, M) \to H^n(C(M)/C^0(M)) \to H_*^{n+1}(L, M) \to \cdots.$$

It is evident that $H^0(C(M)/C^0(M)) = (0)$, whence we conclude that the canonical homomorphism $H_*^1(L, M) \to H^1(L, M)$ is an isomorphism into (a fact which we shall verify directly in the next section). As a by-product of other results, we shall also obtain (in section 3) a simple description of $H^1(C(M)/C^0(M))$ and some information about $H^2(C(M)/C^0(M))$.

2. Extensions of modules. Let L be a Lie algebra over F, and let K and Q be L-modules. An extension of K by Q is a pair (E, ϕ), where E is an L-module containing K, and ϕ is an L-homomorphism of E onto Q whose kernel coincides with K. Two such extensions (E, ϕ) and (E', ϕ') are said to be equivalent if there is an L-isomorphism α of E onto E' which leaves the elements of K fixed and satisfies the relation $\phi'\alpha = \phi$. The equivalence classes of the extensions of K by Q constitute a vector space $\mathrm{ext}(K, Q)$. The definition of the vector space structure of $\mathrm{ext}(K, Q)$ is exactly the same as the one given (for the case of associative algebras) on p. 924 of [4].

Denote by M the F-space of all F-linear maps of Q into K. We make M into an L-module by setting, for $m \in M$, $x \in L$, $q \in Q$, $(x \cdot m)(q) = x \cdot m(q) - m(x \cdot q)$. It is essentially well known that $\mathrm{ext}(K, Q)$ is isomorphic with the ordinary 1-dimensional cohomology group $H^1(L, M)$, regarded in the natural fashion as a vector space over F. We recall only the way in which one attaches a certain cohomology class to each extension (E, ϕ) of K by Q. We can find a linear map ψ of Q into E such that $\phi\psi$ is the identity map on Q. Then we obtain a linear map f of L into M by setting $f(x)(q) = x \cdot \psi(q) - \psi(x \cdot q)$. One verifies directly that f is a Lie 1-cocycle for L in M, and that its cohomology class depends only on the equivalence class of (E, ϕ). It follows immediately from the definition of the vector space structure in $\mathrm{ext}(K, Q)$ that this yields a homomorphism of the F-space $\mathrm{ext}(K, Q)$ into the F-space $H^1(L, M)$. The fact that this is actually an isomorphism onto is proved in exactly the same way as Theorem 1.2 of [4].

Now let us consider the analogous situation in the case where L is a restricted Lie algebra, and K and Q are restricted L-modules. An extension (E, ϕ) of K by Q is then called a restricted extension if E is a restricted L-module. The equivalence classes of the restricted extensions of K by Q constitute a subspace $\mathrm{ext}_*(K, Q)$ of $\mathrm{ext}(K, Q)$. Next we observe that our module M is now a restricted L-module. In fact, by a familiar computation, using that for commuting maps u and v (here, u and v are the endomorphisms of M consisting in composition of an element of M with the x-operator on K or Q, respectively) one has $(u - v)^p = u^p - v^p$, we find that $(x^p \cdot m)(q) = x^p \cdot m(q) - m(x^p \cdot q)$, whence, since K and Q are restricted, $x^p \cdot m = x^{[p]} \cdot m$, which proves our claim.

Let (E, ϕ) be a restricted extension of K by Q, and let f be a Lie 1-cocycle derived from (E, ϕ) as above; i. e., with an evident extension of the notation used before, $f(x) = x \cdot \psi$, where ψ is a linear map of Q into E which is inverse to ϕ. We have $x^{p-1} \cdot f(x) = x^p \cdot \psi$, and since E and Q are

restricted, it follows by the argument just above that $x^{p-1} \cdot f(x) = f(x^{[p]})$, for all $x \in L$. Conversely, let f be any Lie 1-cocycle for L in M such that $x^{p-1} \cdot f(x) = f(x^{[p]})$, for all $x \in L$. We construct a corresponding extension (E, ϕ) of K by Q as follows. The underlying vector space of E is the direct sum $Q + K$, $\phi(q + k) = q$, and the L-operators are defined by $x \cdot (q + k) = x \cdot q + x \cdot k + f(x)(q)$. Then the equivalence class of (E, ϕ) is indeed the element of $\mathrm{ext}(K, Q)$ which corresponds to the cohomology class determined by f. Futhermore, a direct computation shows that

$$x^p \cdot (q + k) = x^{[p]} \cdot q + x^{[p]} \cdot k + \sum_{j=0}^{p-1} x^j \cdot f(x)(x^{p-1-j} \cdot q).$$

If we apply the identity $\sum_{j=0}^{p-1} u^j v u^{p-1-j} = D_u{}^{p-1}(v)$ to the sum on the right, we find that it is equal to $(x^{p-1} \cdot f(x))(q) = f(x^{[p]})(q)$. Hence we conclude that (E, ϕ) is a restricted extension.

In order to interpret this result in the framework of cohomology theory, we consider the canonical homomorphism of $H_*{}^1(L, M)$ into $H^1(L, M)$. We know that it is induced by the cochain map $g \to g^0$ which we have defined at the end of section1. This cochain map is evidently an isomorphism into, and in dimension 0 it is simply the identity map of M onto itself. It follows that the induced homomorphism maps $H_*{}^1(L, M)$ isomorphically into $H^1(L, M)$. Accordingly, we shall identify $H_*{}^1(L, M)$ with the corresponding subspace of $H^1(L, M)$. There remains to decribe this subspace in terms of Lie 1-cocycles. The result is the following:

THEOREM 2. 1. *The canonical homomorphism of $H_*(L, M)$ into $H(L, M)$, where M is any restricted L-module, maps $H_*{}^1(L, M)$ isomorphically onto that subspace of $H^1(L, M)$ whose elements are represented by the Lie 1-cocycles f which satisfy the relation $x^{p-1} \cdot f(x) = f(x^{[p]})$.*

Proof. Let f be a Lie 1-cocycle whose cohomology class belongs to $H_*{}^1(L, M)$. Then there is an associative 1-cocycle g, defined on $U_L{}^+$, such that the cohomology class of g^0 coincides with that of f. This means that there is an element $m \in M$ such that $f(x) = x \cdot m + g(x)$, for all $x \in L$. Hence $x^{p-1} \cdot f(x) = x^p \cdot m + x^{p-1} \cdot g(x) = x^{[p]} \cdot m + g(x^p) = f(x^{[p]})$, since $x^p = x^{[p]}$, in U_L. Conversely, suppose that f satisfies the relation of the theorem. By the equivalence of the associative cochain complex with the Lie cochain complex it follows that f is the restriction to L of an associative cocycle h defined on $R_L{}^+$. We have then, for $r \in R_L$ and $x \in L$,

$$h(r(x^p - x^{[p]})) = rx^{p-1} \cdot h(x) - r \cdot h(x^{[p]}) = r \cdot (x^{p-1} \cdot f(x) - f(x^{[p]})) = 0.$$

Thus, h vanishes on P_L, and is therefore the natural image g^0 of an associative cocycle g defined on U_L^+. Clearly, the cohomology class of g is mapped into that of f by the canonical map of $H_*^1(L, M)$ into $H^1(L, M)$. This completes the proof.

We may now state our result concerning module extensions as follows:

THEOREM 2.2. *Let K and Q be restricted L-modules, and let M be the restricted L-module consisting of the linear maps of Q into K. Then the canonical isomorphism of* $\mathrm{ext}(K, Q)$ *onto* $H^1(L, M)$ *maps* $\mathrm{ext}_*(K, Q)$ *onto* $H_*^1(L, M)$.

3. Extensions of Lie algebras. Let M be an abelian Lie algebra, L an arbitrary Lie algebra. An extension of M by L is a pair (E, ϕ), where E is a Lie algebra containing M as an ideal, and ϕ is a homomorphism of E onto L with kernel M. This situation defines on M the structure of an L-module, with L operating on M via E, in the natural fashion. An automorphism of such an extension is an isomorphism α of E onto E which leaves the elements of M fixed and satisfies the relation $\phi\alpha = \phi$. Then the element $\alpha(e) - e$, with $e \in E$, depends only on $\phi(e)$, and we may write $\alpha(e) = e + f(\phi(e))$, where f is easily seen to be a Lie 1-cocycle for L in M. Conversely, every 1-cocycle gives rise to an automorphism of (E, ϕ) which may be defined by the last relation. Thus the automorphisms of (E, ϕ) constitute a vector space isomorphic with the space $Z^1(L, M)$ of the Lie 1-cocycles for L in M. In particular, if $m \in M$, we obtain a corresponding automorphism $e \to e + [e, m]$, which we call the pseudo-inner automorphism effected by m. Clearly, this gives a rather superficial interpretation of $H^1(L, M)$ as the space of all automorphisms of (E, ϕ), modulo the subspace of the pseudo-inner automorphisms effected by the elements of M. Our purpose in bringing this up is that we wish to illustrate, in the simplest case, a certain difficulty which arises in treating restricted extensions of Lie algebras. Now we assume that M, L, and E are restricted Lie algebras, and that ϕ is a "restricted" homomorphism of E onto L, i. e., $\phi(e^{[p]}) = (\phi(e))^{[p]}$, for every $e \in E$. The pair (E, ϕ) is then called a restricted extension of M by L. It is clear that M is a restricted L-module in this case. It is here appropriate to study the restricted automorphisms α of (E, ϕ). If, as above, we write $\alpha(e) = e + f(\phi(e))$, the condition that α be restricted is easily seen to amount to the condition that $x^{p-1} \cdot f(x) = f(x^{[p]}) - (f(x))^{[p]}$, for all $x \in L$. In fact, for $e \in E$ and $m \in M$, condition (3) on the p-map gives $(m + e)^{[p]} = m^{[p]} + e^{[p]} + e^{p-1} \cdot m$, as is seen from our determination within

E of the term $\sigma(m, e)$. Our assertion concerning f follows at once from this. This result suggests that, in order to obtain interpretations for the restricted cohomology groups in connection with Lie algebra extensions, we should consider restricted extensions in which the kernel M is "strongly abelian," in the sense that $M^{[p]} = (0)$, as well as $[M, M] = (0)$. If M is strongly abelian the above shows that the canonical image of $H_*^1(L, M)$ is precisely that subspace of $H^1(L, M)$ which corresponds to the restricted automorphisms of (E, ϕ).

In general, two restricted extensions (E, ϕ) and (E', ϕ') of M by L will be called *similar* if there exists an ordinary Lie algebra isomorphism γ of E onto E' which leaves the elements of M fixed and for which $\phi'\gamma = \phi$. If, furthermore, γ can be so chosen that $\gamma(e^{[p]}) = (\gamma(e))^{[p]}$, for all $e \in E$, we shall say that (E, ϕ) and (E', ϕ') are *equivalent*.

We place ourselves in the situation where we are given an L-module M, and we consider the set $\mathrm{ext}(M, L)$ of the similarity classes of those extensions of M by L which induce the given L-module structure in M. As is well known, $\mathrm{ext}(M, L)$ carries the structure of a vector space over the base field F, and there is a canonical isomorphism of $\mathrm{ext}(M, L)$ onto $H^2(L, M)$.[8] If L is a restricted Lie algebra and M is a restricted L-module we can consider the subset $\mathrm{ext}_0(M, L)$ of $\mathrm{ext}(M, L)$ which consists of the similarity classes of the restricted extensions. The relation of similarity is so weak that, in considering $\mathrm{ext}_0(M, L)$, the actual p-maps in the extended algebras have little influence, although the requirement that they should exist cannot be neglected. The resulting simplification is expressed by the following lemma.

LEMMA 3.1. *Every restricted extension of M by L is similar to one in which M is strongly abelian.*

Proof. Let (E, ϕ) be a restricted extension of M by L. It follows at once from condition (1) on the p-map that $M^{[p]}$ must be contained in the center of E. If we denote by M^L the subspace of M consisting of all elements that are annihilated by L in the given L-module structure of M, this can be expressed by the inclusion $M^{[p]} \subset M^L$. The restriction to M of the p-map in E is a "p-semilinear" map of M into M^L, in the sense that we have, for $\alpha_i \in F$ and $m_i \in L$, $(\alpha_1 m_1 + \alpha_2 m_2)^{[p]} = \alpha_1^p m_1^{[p]} + \alpha_2^p m_2^{[p]}$. This is seen at once from conditions (2) and (3) on p-maps, noting that M is abelian. We can evidently extend this map to a p-semilinear map k of E into M^L. It is then

[8] The constructions which are needed in this connection will appear in our proofs below. The 0-element of $\mathrm{ext}(M, L)$ is the class of the trivial extension which contains L as a subalgebra.

Thus, h vanishes on P_L, and is therefore the natural image g^0 of an associative cocycle g defined on U_L^+. Clearly, the cohomology class of g is mapped into that of f by the canonical map of $H_*^1(L, M)$ into $H^1(L, M)$. This completes the proof.

We may now state our result concerning module extensions as follows:

THEOREM 2.2. *Let K and Q be restricted L-modules, and let M be the restricted L-module consisting of the linear maps of Q into K. Then the canonical isomorphism of* $\mathrm{ext}(K, Q)$ *onto* $H^1(L, M)$ *maps* $\mathrm{ext}_*(K, Q)$ *onto* $H_*^1(L, M)$.

3. Extensions of Lie algebras. Let M be an abelian Lie algebra, L an arbitrary Lie algebra. An extension of M by L is a pair (E, ϕ), where E is a Lie algebra containing M as an ideal, and ϕ is a homomorphism of E onto L with kernel M. This situation defines on M the structure of an L-module, with L operating on M via E, in the natural fashion. An automorphism of such an extension is an isomorphism α of E onto E which leaves the elements of M fixed and satisfies the relation $\phi\alpha = \phi$. Then the element $\alpha(e) - e$, with $e \in E$, depends only on $\phi(e)$, and we may write $\alpha(e) = e + f(\phi(e))$, where f is easily seen to be a Lie 1-cocycle for L in M. Conversely, every 1-cocycle gives rise to an automorphism of (E, ϕ) which may be defined by the last relation. Thus the automorphisms of (E, ϕ) constitute a vector space isomorphic with the space $Z^1(L, M)$ of the Lie 1-cocycles for L in M. In particular, if $m \in M$, we obtain a corresponding automorphism $e \to e + [e, m]$, which we call the pseudo-inner automorphism effected by m. Clearly, this gives a rather superficial interpretation of $H^1(L, M)$ as the space of all automorphisms of (E, ϕ), modulo the subspace of the pseudo-inner automorphisms effected by the elements of M. Our purpose in bringing this up is that we wish to illustrate, in the simplest case, a certain difficulty which arises in treating restricted extensions of Lie algebras. Now we assume that M, L, and E are restricted Lie algebras, and that ϕ is a "restricted" homomorphism of E onto L, i. e., $\phi(e^{[p]}) = (\phi(e))^{[p]}$, for every $e \in E$. The pair (E, ϕ) is then called a restricted extension of M by L. It is clear that M is a restricted L-module in this case. It is here appropriate to study the restricted automorphisms α of (E, ϕ). If, as above, we write $\alpha(e) = e + f(\phi(e))$, the condition that α be restricted is easily seen to amount to the condition that $x^{p-1} \cdot f(x) = f(x^{[p]}) - (f(x))^{[p]}$, for all $x \in L$. In fact, for $e \in E$ and $m \in M$, condition (3) on the p-map gives $(m + e)^{[p]} = m^{[p]} + e^{[p]} + e^{p-1} \cdot m$, as is seen from our determination within

E of the term $\sigma(m, e)$. Our assertion concerning f follows at once from this. This result suggests that, in order to obtain interpretations for the restricted cohomology groups in connection with Lie algebra extensions, we should consider restricted extensions in which the kernel M is "strongly abelian," in the sense that $M^{[p]} = (0)$, as well as $[M, M] = (0)$. If M is strongly abelian the above shows that the canonical image of $H_*^1(L, M)$ is precisely that subspace of $H^1(L, M)$ which corresponds to the restricted automorphisms of (E, ϕ).

In general, two restricted extensions (E, ϕ) and (E', ϕ') of M by L will be called *similar* if there exists an ordinary Lie algebra isomorphism γ of E onto E' which leaves the elements of M fixed and for which $\phi'\gamma = \phi$. If, furthermore, γ can be so chosen that $\gamma(e^{[p]}) = (\gamma(e))^{[p]}$, for all $e \in E$, we shall say that (E, ϕ) and (E', ϕ') are *equivalent*.

We place ourselves in the situation where we are given an L-module M, and we consider the set $\mathrm{ext}(M, L)$ of the similarity classes of those extensions of M by L which induce the given L-module structure in M. As is well known, $\mathrm{ext}(M, L)$ carries the structure of a vector space over the base field F, and there is a canonical isomorphism of $\mathrm{ext}(M, L)$ onto $H^2(L, M)$.[8] If L is a restricted Lie algebra and M is a restricted L-module we can consider the subset $\mathrm{ext}_0(M, L)$ of $\mathrm{ext}(M, L)$ which consists of the similarity classes of the restricted extensions. The relation of similarity is so weak that, in considering $\mathrm{ext}_0(M, L)$, the actual p-maps in the extended algebras have little influence, although the requirement that they should exist cannot be neglected. The resulting simplification is expressed by the following lemma.

LEMMA 3.1. *Every restricted extension of M by L is similar to one in which M is strongly abelian.*

Proof. Let (E, ϕ) be a restricted extension of M by L. It follows at once from condition (1) on the p-map that $M^{[p]}$ must be contained in the center of E. If we denote by M^L the subspace of M consisting of all elements that are annihilated by L in the given L-module structure of M, this can be expressed by the inclusion $M^{[p]} \subset M^L$. The restriction to M of the p-map in E is a "p-semilinear" map of M into M^L, in the sense that we have, for $\alpha_i \in F$ and $m_i \in L$, $(\alpha_1 m_1 + \alpha_2 m_2)^{[p]} = \alpha_1^p m_1^{[p]} + \alpha_2^p m_2^{[p]}$. This is seen at once from conditions (2) and (3) on p-maps, noting that M is abelian. We can evidently extend this map to a p-semilinear map k of E into M^L. It is then

[8] The constructions which are needed in this connection will appear in our proofs below. The 0-element of $\mathrm{ext}(M, L)$ is the class of the trivial extension which contains L as a subalgebra.

easily seen that the map $e \to e^{(p)} = e^{[p]} - k(e)$ still satisfies the conditions for a p-map. If E is equipped with this new p-map, we have a restricted extension of M by L which is similar to the given extension by the identity map of E onto E, and in which M is strongly abelian.

A first consequence of this result is that $\mathrm{ext}_0(M, L)$ is a subspace of $\mathrm{ext}(M, L)$. In order to prove this, we must recall the definition of the vector space structure in $\mathrm{ext}(M, L)$. The subtraction in $\mathrm{ext}(M, L)$ is the map which is induced by the following composition of extensions (E, ϕ) and (E', ϕ'). Let D denote the subalgebra of the direct sum of E and E' which consists of all elements (e, e') in which $\phi(e) = \phi'(e')$. Let J be the ideal of D consisting of the elements (m, m), with $m \in M$. In D/J, we identify $(M, M)/J$ with M by means of the homomorphism $(m_1, m_2) \to m_1 - m_2$ whose kernel is exactly J. The homomorphism $(e, e') \to \phi(e)$ induces a homomorphism ψ of D/J onto L, and $(D/J, \psi)$ is the extension of M by L which represents the difference of the similarity classes of (E, ϕ) and (E', ϕ'). Now if (E, ϕ) and (E', ϕ') are restricted extensions, we can define a p-map in D by setting $(e, e')^{[p]} = (e^{[p]}, e'^{[p]})$. Since this maps J into itself, it induces a p-map in D/J, with which $(D/J, \psi)$ constitutes a restricted extension of M by L. Hence $\mathrm{ext}_0(M, L)$ is a subgroup of $\mathrm{ext}(M, L)$.

The scalar multiple of the similarity class of (E, ϕ) by an element $\alpha \in F$ is the similarity class of the extension defined as follows. First construct a Lie algebra (E, M) whose space is the direct sum of E and M, and where the commutation is defined by the formula

$$[(e_1, m_1), (e_2, m_2)] = ([e_1, e_2], e_1 \cdot m_2 - e_2 \cdot m_1).$$

Let J_α be the ideal of (E, M) which consists of the elements of the form $(m, -\alpha m)$, with $m \in M$. In $(E, M)/J_\alpha$, identify $(M, M)/J_\alpha$ with M by the homomorphism $(m_1, m_2) \to \alpha m_1 + m_2$ whose kernel is J_α. Let ψ be the homomorphism of $(E, M)/J_\alpha$ onto L which is induced by the homomorphism $(e, m) \to \phi(e)$. Then the similarity class of $((E, M)/J_\alpha, \psi)$ is the desired α-multiple of the similarity class of (E, ϕ).

We wish to show that multiplication by α maps $\mathrm{ext}_0(M, L)$ into itself. By Lemma 3.1, every element of $\mathrm{ext}_0(M, L)$ is represented by a restricted extension (E, ϕ) in which M is strongly abelian. In that case, the map $(e, m) \to (e, m)^{[p]} = (e^{[p]}, e^{p-1} \cdot m)$ annihilates (M, M) and therefore induces a p-map in $(E, M)/J_\alpha$ (which annihilates M). With this p-map, $((E, M)/J_\alpha, \psi)$ is a restricted extension of M by L (in which M is strongly abelian). This completes the proof that $\mathrm{ext}_0(M, L)$ is a subspace of $\mathrm{ext}(M, L)$.

Now denote by $H_r^2(L, M)$ the image of $\mathrm{ext}_0(M, L)$ in $H^2(L, M)$, by the

canonical isomorphism of $\mathrm{ext}(M, L)$ onto $H^2(L, M)$. We wish to characterize $H_0^2(L, M)$ in terms of cohomology theory. Let f be any Lie 2-cocycle for L in M. We consider an extension of M by L whose similarity class corresponds to the cohomology class c of f, by the canonical isomorphism. Such an extension is the following. The underlying space of the extended algebra is the direct sum of L and M, and the commutation is defined by the formula $[(x, m), (x_1, m_1)] = ([x, x_1], x \cdot m_1 - x_1 \cdot m + f(x, x_1))$. The homomorphism onto L is the map $(x, m) \to x$. It is easy to verify that we have

$$D_{(x,0)^p}(x_1, 0) = ([x^{[p]}, x_1], k_x(x_1)), \text{ where } k_x(x_1) = \sum_{i=0}^{p-1} x^i \cdot f(x, D_x^{p-1-i}(x_1)).$$

If we apply $D_{(x,0)^p}$ to the commutator $[(x_1, 0), (x_2, 0)]$ and express the fact that $D_{(x,0)^p}$ is a derivation, we obtain

$$k_x([x_1, x_2]) + x^{[p]} \cdot f(x_1, x_2) = x_1 \cdot k_x(x_2) - x_2 \cdot k_x(x_1) + f(x_1, [x^{[p]}, x_2]) + f([x^{[p]}, x_1], x_2).$$

If we write $f_y(x) = f(x, y)$, this relation means precisely that, for each fixed $x \in L$, $k_x + f_{x^{[p]}}$ is a Lie 1-cocycle for L in M.

If f is a coboundary, δg, we obtain

$$k_x(x_1) = \sum_{i=0}^{p-1} x^i \cdot (x \cdot g(D_x^{p-1-i}(x_1)) - D_x^{p-1-i}(x_1) \cdot g(x) - g(D_x^{p-i}(x_1)))$$
$$= x^{[p]} \cdot g(x_1) - g([x^{[p]}, x_1]) - \sum_{i=0}^{p-1} x^i D_x^{p-1-i}(x_1) \cdot g(x).$$

Now the identity (valid in any commutative ring)

$$u^k - (u - v)^k = v \sum_{i=0}^{k-1} u^i (u - v)^{k-1-i}$$

shows that

$$\sum_{i=0}^{k-1} u^i (u - v)^{k-1-i} = \sum_{j=0}^{k-1} (-1)^j \binom{k}{j+1} u^{k-1-j} v^j.$$

If we specialize this to the case where u is the left multiplication by x in R_L and v is the right multiplication by x in R_L, we find that

$$\sum_{i=0}^{k-1} x^i D_x^{k-1-i}(x_1) = \sum_{j=0}^{k-1} (-1)^j \binom{k}{j+1} x^{k-1-j} x_1 x^j. \tag{1}$$

Set $k = p$ in this relation and substitute the result in the above expression for $k_x(x_1)$. This gives

$$k_x(x_1) = x^{[p]} \cdot g(x_1) - g([x^{[p]}, x_1]) - x_1 x^{p-1} \cdot g(x),$$

whence

$$k_x(x_1) + f(x_1, x^{[p]}) = x_1 \cdot (g(x^{[p]}) - x^{p-1} \cdot g(x)).$$

It follows that the cohomology class of $k_x + f_{x^{[p]}}$ depends only on the cohomology class c of f. We shall denote it by $c(x)$.

If $c \in H_0^2(L, M)$, we can define a p-map in our extended algebra so as to obtain a restricted extension of M by L. Write $(x, 0)^{[p]} = (x^{[p]}, \rho(x))$. Then, since $D_{(x,0)^p} = D_{(x,0)^{[p]}}$, we must have $k_x(x_1) = f(x^{[p]}, x_1) - x_1 \cdot \rho(x)$, i. e., $k_x + f_{x^{[p]}} = \delta(-\rho(x))$, so that $c(x) = 0$.

We wish to prove, conversely, that if $c(x) = 0$, for all $x \in L$, then $c \in H_0^2(L, M)$. We can find an associative 2-cocycle g for R_{L^+} in M which represents the given cohomology class $c \in H^2(L, M)$. If

$$f(x, y) = g(x, y) - g(y, x)$$

then f is a Lie 2-cocycle belonging to c. If we express $k_x(x_1)$ in terms of values of g we obtain, after some obvious simplifications resulting from the fact that g is a cocycle,

$$k_x(x_1) = g(x^p, x_1) - \sum_{i=0}^{p-1} g(x^i D_x^{p-1-i}(x_1), x).$$

By the above identity (1), this becomes

$$\begin{aligned} k_x(x_1) &= g(x^p, x_1) - g(x_1 x^{p-1}, x) \\ &= g(x^p, x_1) - g(x_1, x^p) - x_1 \cdot g(x^{p-1}, x). \end{aligned}$$

Hence $c(x)$ has the representative 1-cocycle g'_x, where

$$g'_x(x_1) = g(x^p - x^{[p]}, x_1) - g(x_1, x^p - x^{[p]}).$$

This expression for g'_x shows at once that the map $x \to c(x)$ is a p-semilinear map of L into $H^1(L, M)$, and that if $c(x) = 0$, for all $x \in L$, we can find a p-semilinear map σ of L into M such that we have identically $g'_x(x_1) = x_1 \cdot \sigma(x)$. Now we define a p-map in our extended algebra by setting $(x, m)^{[p]} = (x^{[p]}, x^{p-1} \cdot m + g(x^{p-1}, x) - \sigma(x))$. Condition (2) for a p-map is evidently satisfied. The verification of conditions (1) and (3) requires some computation. By induction on k, one finds that

$$\begin{aligned} &D_{(x,m)}{}^k(x_1, m_1) \\ &\quad = (D_x{}^k(x_1), x^k \cdot m_1 - \sum_{i=0}^{k-1} x^i D_x{}^{k-1-i}(x_1) \cdot m + \sum_{i=0}^{k-1} x^i \cdot f(x, D_x{}^{k-1-i}(x_1))). \end{aligned}$$

If we write the terms of the last sum on the right in terms of values of g, we find that it is equal to $g(x^k, x_1) - \sum_{i=0}^{k-1} g(x^i D_x{}^{k-1-i}(x_1), x)$. If we take $k = p$ and use the above identity (1) we obtain

$$D_{(x,m)^p}(x_1, m_1) = ([x^{[p]}, x_1], x^{[p]} \cdot m_1 - x_1 x^{p-1} \cdot m + g(x^p, x_1) - g(x_1 x^{p-1}, x)),$$

which is easily seen to be equal to $[(x, m)^{[p]}, (x_1, m_1)]$.[9] Thus our map satisfies condition (1).

If, in our identity (1), we put $k = p - 1$ and note that $\binom{p-1}{s}$ is congruent to $(-1)^s$ mod. p, we find that $\sum_{i=0}^{p-2} x^i D_x{}^{p-2-i}(x_1) = -\sum_{j=0}^{p-2} x^{p-2-j} x_1 x^j$. Now we operate in the polynomial ring $R_E[t]$ as in our discussion of the p-map, in section 1. Rewriting our above result for $D_{(x,m)}{}^k(x_1, m_1)$, with $k = p - 1$ and with $(x_1, m_1)t + (x, m)$ in the place of (x, m), and using the last identity with $x_1 t + x$ in the place of x, we obtain

$$D_{(x_1,m_1)t+(x,m)}{}^{p-1}(x_1, m_1) = (D_{x_1 t+x}{}^{p-1}(x_1), q(t)),$$

where $q(t)$ is the polynomial in t with coefficients in M which is obtained by expanding the expression

$$(x_1 t + x)^{p-1} \cdot m_1 + \sum_{j=0}^{p-2} (x_1 t + x)^{p-2-j} x_1 (x_1 t + x)^j \cdot (m_1 t + m)$$
$$+ g((x_1 t + x)^{p-1}, x_1) + \sum_{j=0}^{p-2} g((x_1 t + x)^{p-2-j} x_1 (x_1 t + x)^j, (x_1 t + x)).$$

It is seen at once from this expression that $q(t)$ is the formal derivative with respect to t of the polynomial obtained by expanding the expression $(x_1 t + x)^{p-1} \cdot (m_1 t + m) + g((x_1 t + x)^{p-1}, (x_1 t + x))$. Note that this implies that $q(t)$ is actually of degree less than $p - 1$. If we formally integrate $q(t)$ with respect to t, regarding it as a polynomial of degree $p - 2$, and then put $t = 1$ we obtain therefore

$$(x_1 + x)^{p-1} \cdot (m_1 + m) + g((x_1 + x)^{p-1}, (x_1 + x))$$
$$- x_1{}^{p-1} \cdot m_1 - g(x_1{}^{p-1}, x_1) - x^{p-1} \cdot m - g(x^{p-1}, x);$$

namely those terms of the last expression in which the exponent of t is greater than 0 and less than p. On the other hand, the component in M of $((x, m) + (x_1, m_1))^{[p]}$ is

$$(x_1 + x)^{p-1} \cdot (m_1 + m) + g((x_1 + x)^{p-1}, (x_1 + x)) - \sigma(x_1 + x).$$

Hence we see that the formal integral of $q(t)$ is precisely the component in M of the element $((x, m) + (x_1, m_1))^{[p]} - (x, m)^{[p]} - (x_1, m_1)^{[p]}$. By what we have seen in our discussion of section 1, this means that condition (3) is satisfied by our p-map. Hence we have the following result.

THEOREM 3.1. *Let L be a restricted Lie algebra, M a restricted L-*

[9] It is here that the relation $g'_x(x_1) = x_1 \cdot \sigma(x)$ is used.

module. Denote by $S(L, H^1(L, M))$ *the space of all p-semilinear maps of* L *into* $H^1(L, M)$. *For an associative 2-cocycle* g, *let* g' *be the function defined by* $g'_x(y) = g(x^p - x^{[p]}, y) - g(y, x^p - x^{[p]})$. *Then the map* $g \to g'$ *induces an F-linear map of* $H^2(L, M)$ *into* $S(L, H^1(L, M))$ *whose kernel is precisely the subspace* $H_0^2(L, M)$ *that corresponds to the restricted extensions of* M *by* L.

We shall now look at our situation more closely from the following point of view. We suppose that we are given a restricted L-module M, together with a p-semilinear map of M into M^L. Then M has also the structure of a restricted abelian Lie algebra. We shall refer to these data by saying that M is given as a restricted abelian Lie algebra on which L operates. The object to be investigated is the set $\mathrm{ext}_*(M, L)$ of the equivalence classes of those restricted extensions of M by L which induce the given L-module structure and p-map on M. If we assign to each such equivalence class the similarity class to which it belongs we obtain what we shall call the canonical map of $\mathrm{ext}_*(M, L)$ into $\mathrm{ext}(M, L)$. It is clear from Lemma 3.1 that, no matter what the given p-map in M actually is, the canonical map sends $\mathrm{ext}_*(M, L)$ *onto* $\mathrm{ext}_0(M, L)$. Furthermore, it is clear from our discussion of the vector space structure in $\mathrm{ext}_0(M, L)$ that $\mathrm{ext}_*(M, L)$ can be given the structure of an abelian group such that the canonical map onto $\mathrm{ext}_0(M, L)$ is a group homomorphism. In fact, this is done by means of the same composition of extensions which we have used for making $\mathrm{ext}_0(M, L)$ into an abelian group. Furthermore, our earlier discussion shows that, in the case where M is strongly abelian, our previous construction (for $\mathrm{ext}_0(M, L)$) yields also scalar multiples in $\mathrm{ext}_*(M, L)$ so that $\mathrm{ext}_*(M, L)$ becomes a vector space, and the canonical map onto $\mathrm{ext}_0(M, L)$ is an F-linear map. In the general case, we do not obtain a vector space structure in $\mathrm{ext}_*(M, L)$.

Next we wish to describe the partitioning of a similarity class c of restricted extensions of M by L into a set S_c of equivalence classes. Let γ be a similarity isomorphism of a restricted extension (E, ϕ) onto a restricted extension (E', ϕ'). For $e \in E$ and $m \in M$, we have then

$$\gamma((e + m)^{[p]}) = \gamma(e^{[p]} + m^{[p]} + e^{p-1} \cdot m) = \gamma(e^{[p]}) + m^{[p]} + e^{p-1} \cdot m,$$

while

$$(\gamma(e + m))^{[p]} = (\gamma(e) + m)^{[p]} = (\gamma(e))^{[p]} + m^{[p]} + \gamma(e)^{p-1} \cdot m.$$

It is seen from this that $\gamma(e^{[p]}) - (\gamma(e))^{[p]}$ depends only on the coset of e mod. M, so that we may write this difference in the form $g(\phi(e))$, where g is a map of L into M. Our discussion of condition (3) for a p-map shows

that we have here $\gamma(\sigma(e_1, e_2)) = \sigma(\gamma(e_1), \gamma(e_2))$. It follows from this that g is p-semilinear. Furthermore, for $x = \phi(z) \in L$, we have

$$x \cdot g(\phi(e)) = \gamma(z) \cdot (\gamma(e^{[p]}) - (\gamma(e))^{[p]}) = \gamma([z, e^{[p]}]) + D_{\gamma(e)^p}(\gamma(z))$$
$$= \gamma([z, e^{[p]}) + \gamma(D_e{}^p(z)) = 0,$$

so that g actually maps L into M^L. Thus g belongs to the space $S(L, M^L)$ of the p-semilinear maps of L into M^L. Now denote by E_g the restricted Lie algebra which is obtained from E by changing the p-map to

$$e \to e^{(p)} = e^{[p]} - g(\phi(e)).$$

Then (E_g, ϕ) is a restricted extension of M by L and γ is an equivalence isomorphism of (E_g, ϕ) onto (E', ϕ'). Conversely, if g is any p-semilinear map of L into M^L, then (E_g, ϕ) is a restricted extension of M by L which is similiar to (E, ϕ).

With every $g \in S(L, M^L)$ we associate the transformation g^* on $\mathrm{ext}_*(M, L)$ which is induced by the map $(E, \phi) \to (E_g, \phi)$. We have just seen that $g^*(S_c) \subset S_c$, and that $S(L, M^L)$ operates transitively on each S_c. The transformations g^* are actually group translations in $\mathrm{ext}_*(M, L)$. In fact, let s_0 denote the 0-element of $\mathrm{ext}_*(M, L)$, i. e., the equivalence class of the trivial extension. Then, for any $s \in \mathrm{ext}_*(M, L)$, we have $g^*(s) = s + g^*(s_0)$, as is easily seen from the definitions. Furthermore, the maps $g \to g^*$ and $g \to g^*(s_0)$ are group homomorphisms, as a glance at the definitions will show.

Next we shall determine the kernel of the map $g \to g^*$. Suppose that ζ is an equivalence isomorphism of (E_{g_1}, ϕ) onto (E_{g_2}, ϕ). Denote the p-maps in E_{g_1} and E_{g_2} by $e \to e^{\{p\}}$ and $e \to e^{(p)}$, respectively. We write $\zeta(e) = e + h(\phi(e))$, where h is a Lie 1-cocycle for L in M, as we did when we studied the automorphism of an extension. If we write out the relation $\zeta(e^{\{p\}}) = (\zeta(e))^{(p)}$ we find that we must have, for every $x \in L$,

$$g_2(x) - g_1(x) = x^{p-1} \cdot h(x) - h(x^{[p]}) + (h(x))^{[p]}.$$

We write this as $g_2 - g_1 = h'$. Conversely, let h be any Lie 1-cocycle for L in M. We know that h is the restriction to L of an associative cocycle f defined on $R_L{}^+$. Hence

$$h'(x) = x^{p-1} \cdot f(x) - f(x^{[p]}) + (h(x))^{[p]} = f(x^p - x^{[p]}) + (h(x))^{[p]},$$

whence we see that h' is p-semilinear and takes its values in M^L. Thus we have a corresponding extension $(E_{h'}, \phi)$, and the map $e \to \zeta(e) = e + h(\phi(e))$ is an equivalence isomorphism of (E, ϕ) onto $(E_{h'}, \phi)$. Hence the kernel

of the map $g \to g^*$ is precisely the image in $S(L, M^L)$ of the space $Z^1(L, M)$ of the Lie 1-cocycles for L in M, under the above linear map $h \to h'$.

By the above, the map $g \to g^*(s_0)$ of $S(L, M^L)$ into $\mathrm{ext}_*(M, L)$ has exactly the same kernel, while its image is precisely the kernel of the canonical homomorphism of $\mathrm{ext}_*(M, L)$ onto $\mathrm{ext}_0(M, L)$. Thus we have the following result.

THEOREM 3.2. *Let M be an abelian restricted Lie algebra on which the restricted Lie algebra L operates. Then the linear map $h \to h'$ of $Z^1(L, M)$ into $S(L, M^L)$, where $h'(x) = x^{p-1} \cdot h(x) - h(x^{[p]}) + (h(x))^{[p]}$, the homomorphism $g \to g^*(s_0)$ of $S(L, M^L)$ into $\mathrm{ext}_*(M, L)$, and the canonical homomorphism of $\mathrm{ext}_*(M, L)$ onto $\mathrm{ext}_0(M, L)$ constitute an exact sequence*

$$Z^1(L, M) \to S(L, M^L) \to \mathrm{ext}_*(M, L) \to \mathrm{ext}_0(M, L) \to (0).$$

From now on we suppose that M is strongly abelian and connect $\mathrm{ext}_*(M, L)$ with $H_*{}^2(L, M)$. Suppose that an associative 2-cocycle g for $U_L{}^+$ in M is given. We construct a corresponding restricted extension of M by L as follows. The space of the extended Lie algebra, E_g, is the direct sum of L and M. The commutation is defined by the formula

$$[(x_1, m_1), (x_2, m_2)] = ([x_1, x_2], x_1 \cdot m_2 - x_2 \cdot m_1 + g(x_1, x_2) - g(x_2, x_1)).$$

The homomorphism ϕ_g of E_g onto L is defined by $\phi_g(x, m) = x$. Finally, we define a p-map in E_g by setting $(x, m)^{[p]} = (x^{[p]}, x^{p-1} \cdot m + g(x^{p-1}, x))$. In fact, in the course of our proof of Theorem 3.1, we have already seen that this map satisfies all the conditions for a p-map. Next we show that the equivalence class of our restricted extension (E_g, ϕ_g) depends only on the cohomology class, in $H_*{}^2(L, M)$, of g. Let h be any 1-cochain for $U_L{}^+$ in M. For $(x, m) \in E_{g+\delta h}$, set $\alpha(x, m) = (x, m + h(x)) \in E_g$. It is easy to verify as usual that α is a similarity isomorphism of $(E_{g+\delta h}, \phi_{g+\delta h})$ onto (E_g, ϕ_g). Furthermore, we have

$$\begin{aligned}\alpha((x, m)^{[p]}) &= \alpha(x^{[p]}, x^{p-1} \cdot m + g(x^{p-1}, x) + x^{p-1} \cdot h(x) - h(x^p)) \\ &= (x^{[p]}, x^{p-1} \cdot m + g(x^{p-1}, x) + x^{p-1} \cdot h(x)) = (\alpha(x, m))^{[p]},\end{aligned}$$

i. e., α is an equivalence isomorphism.

Thus, the correspondence $g \to (E_g, \phi_g)$ induces a map of $H_*{}^2(L, M)$ into $\mathrm{ext}_*(M, L)$. Now it is not difficult to see that the difference of the equivalence classes of (E_{g_1}, ϕ_{g_1}) and (E_{g_2}, ϕ_{g_2}) is represented by the extension $(E_{g_1-g_2}, \phi_{g_1-g_2})$. Similarly, the scalar multiple by an element $\alpha \in F$ of the equivalence class of (E_g, ϕ_g) is represented by $(E_{\alpha g}, \phi_{\alpha g})$. Hence our map

of $H_*^2(L, M)$ into $\mathrm{ext}_*(M, L)$ is an F-linear homomorphism. We shall see that it is actually an isomorphism.

Suppose that (E_g, ϕ_g) is a trivial extension. This means that there is a restricted isomorphism ψ of L into E_g such that $\phi_g\psi$ is the identity on L. Write $\psi(x) = (x, -h(x))$. Then we have

$$g(x, y) - g(y, x) = x \cdot h(y) - y \cdot h(x) - h([x, y]),$$

and

$$g(x^{p-1}, x) = x^{p-1} \cdot h(x) - h(x^{[p]}).$$

The first relation means that the Lie cocycle corresponding to g is a coboundary. Now this Lie cocycle is also the Lie cocycle which correspond to the associative 2-cocycle g^0 for R_{L^+} in M, and we conclude that g^0 is also a coboundary, i. e., there is a 1-cochain c for R_{L^+} in M such that $g^0(u, v) = u \cdot c(v) - c(uv)$, for all u, v in R_{L^+}. If c' is the restriction of c to L then $c' - h$ is evidently a Lie 1-cocycle, and therefore coincides with the restriction k' to L of an associative 1-cocycle k for R_{L^+} in M. Replacing c by $c - k$, we may therefore assume that $c' = h$. Then we have from the above $g(x^{p-1}, x) = x^{p-1} \cdot c(x) - c(x^{[p]})$, for every $x \in L$. But this means that $c(x^p) = c(x^{[p]})$, i. e., c vanishes on P_L, so that $c = f^0$, with f a 1-cochain for U_{L^+} in M. Clearly, we have $g = \delta f$, and so we conclude that our homomorphism of $H_*^2(L, M)$ into $\mathrm{ext}_*(M, L)$ is an isomorphism. There remains to show that it maps $H_*^2(L, M)$ onto $\mathrm{ext}_*(M, L)$.

Let (E, ϕ) be a restricted extension of M by L. The homomorphism ϕ can be extended uniquely to a homomorphism ϕ' of U_E onto U_L such that $\phi'(1) = 1$. The kernel of ϕ' is the ideal $U_E M$ of U_E which is generated by the elements of M. We claim that the identity map of M onto M can be extended in one and only one way to a U_E-homomorphism of $U_E M$ onto M, each being regarded as a U_E-module in the natural fashion. Clearly, there can be at most one such map. Let (m_i) be a basis for M over F. We claim that if $\sum_i u_i m_i = 0$, with $u_i \in U_E$, then each u_i must actually belong to $U_E M$. In fact, if we multiply our relation by the product of the elements m_i^{p-1}, where i ranges over all the indices which occur in the sum, with the exception of one of them, say $i = 1$, we obtain $u_1 m_1 m_2^{p-1} \cdots m_k^{p-1} = 0$. If we write u_1 as a linear combination of ordered monomials in the elements of an ordered basis of E which contains the m_i we see at once that this last relation implies that $u_1 \in U_E M$, and this suffices to establish our claim. Hence there exists a map γ of $U_E M$ onto M such that $\gamma(\sum_i u_i m_i) = \sum_i \phi'(u_i) \cdot m_i$, and it is evident that γ is a U_E-homomorphism.

Now let ψ be a linear map of L into E which is inverse to ϕ. We can extend ψ to a linear map ψ' of U_L into U_E such that $\phi'\psi'$ is the identity map on U_L. For x and y in $U_L{}^+$, we now set $g(x, y) = \gamma(\psi'(x)\psi'(y) - \psi'(xy))$, noting that $\psi'(x)\psi'(y) - \psi'(xy)$ lies in the kernel of ϕ', i. e., in $U_E M$. Now observe that, for $e \in E$ and $m \in M$, we have $me = em - \phi(e) \cdot m$, whence $\gamma(me) = 0$. Hence we conclude that γ maps $U_E M U_E{}^+$ into (0). Using this, it is easily verified that $x \cdot g(y, z) - g(xy, z) + g(x, yz) = 0$, for all x, y and z in $U_L{}^+$, which means that g is an associative 2-cocycle for $U_L{}^+$ in M.

Now we shall show that (E, ϕ) is equivalent to (E_g, ϕ_g). Define a map $\alpha\colon E_g \to E$ by setting $\alpha(x, m) = \psi(x) + m$. Then we have

$$\begin{aligned}
&\alpha([(x_1, m_1), (x_2, m_2)]) \\
&\quad = \psi([x_1, x_2]) + x_1 \cdot m_2 - x_2 \cdot m_1 + g(x_1, x_2) - g(x_2, x_1) \\
&\quad = \psi([x_1, x_2]) + x_1 \cdot m_2 - x_2 \cdot m_1 + \gamma([\psi(x_1), \psi(x_2)] - \psi([x_1, x_2])) \\
&\quad = [\psi(x_1), \psi(x_2)] + x_1 \cdot m_2 - x_2 \cdot m_1 = [\alpha(x_1, m_1), \alpha(x_2, m_2)],
\end{aligned}$$

whence α is a similarity isomorphism. Furthermore,

$$(\alpha(x, m))^{[p]} = (\psi(x))^{[p]} + x^{p-1} \cdot m,$$

while

$$\alpha((x, m)^{[p]}) = \psi(x^{[p]}) + g(x^{p-1}, x) + x^{p-1} \cdot m.$$

But we have

$$\begin{aligned}
g(x^{p-1}, x) &= \gamma(\psi'(x^{p-1})\psi(x) - \psi'(x^p)) \\
&= \gamma(\psi(x)^p - \psi(x^{[p]})) = (\psi(x))^{[p]} - \psi(x^{[p]}),
\end{aligned}$$

whence $(\alpha(x, m))^{[p]} = \alpha((x, m)^{[p]})$, i. e., α is an equivalence isomorphism.

Hence our map $H_*{}^2(L, M) \to \mathrm{ext}_*(M, L)$ is an isomorphism onto, and our above construction yields its inverse. We have therefore obtained the following result:

THEOREM 3.3. *Let M be a strongly abelian restricted Lie algebra on which the restricted Lie algebra L operates. Then there is a canonical isomorphism of $\mathrm{ext}_*(M, L)$ onto $H_*{}^2(L, M)$ such that the following diagram of canonical F-linear maps is commutative:*

$$\begin{array}{ccc}
\mathrm{ext}_*(M, L) & \to & H_*{}^2(L, M) \\
\downarrow & & \downarrow \\
\mathrm{ext}_0(M, L) & \to & H_0{}^2(L, M).
\end{array}$$

In the case where M is strongly abelian, we can replace $Z^1(L, M)$ by $H^1(L, M)$ in the exact sequence of Theorem 3. 2. Taking account of Theorems 3. 1 and 2. 1 also, we find that, for every restricted L-module M, the following sequence is exact:

$$(0) \to H_*^1(L, M) \to H^1(L, M) \to S(L, M^L)$$
$$\to H_*^2(L, M) \to H^2(L, M) \to S(L, H^1(L, M)).$$

We shall compare this sequence with that obtained at the end of section 1, which involved the relative cohomology groups $H^n(C(M)/C^0(M))$. An element $a \in H^1(C(M)/C^0(M))$ is represented by a cochain $f \in C(M)$ such that $\delta f \in C^0(M)$. Thus f is a linear map of R_L^+ into M such that $u \cdot f(v) - f(uv)$ depends only on the cosets mod. P_L of u and v. Now define a p-semilinear map g of L into M by setting $g(x) = f(x^p - x^{[p]})$. It is seen immediately that g does not depend on our particular choice of the representative f, but only on a. Furthermore, a direct verification shows that $g \in S(L, M^L)$. Hence the correspondence $a \to g$ is actually an F-linear map of $H^1(C(M)/C^0(M))$ into $S(L, M^L)$. We shall denote this map by γ_1. Similarly, we shall define an F-linear map γ_2 of $H^2(C(M)/C^0(M))$ into $S(L, H^1(L, M))$. In fact, an element $a \in H^2(C(M)/C^0(M))$ is represented by a cochain $f \in C(M)$ such that $\delta f \in C^0(M)$, i. e., by a bilinear map f from R_L^+ to M such that $u \cdot f(v, w) - f(uv, w) + f(u, vw)$ depends only on the cosets mod. P_L of u, v, and w. For $x \in L$, we now set $g_x(u) = f(x^p - x^{[p]}, u) - f(u, x^p - x^{[p]})$. Then, as we know already from Theorem 3. 1, the map $x \to g_x$ induces a p-semilinear map of L into $H^1(L, M)$. Furthermore, it is seen immediately that the cohomology class of g_x does not depend on our particular choice of the representative f, but only on a. If we denote this by $\gamma_2(a)$, we have the desired F-linear map γ_2 of $H^2(C(M)/C^0(M))$ into $S(L, H^1(L, M))$. Now let us consider the two exact sequences with the connecting maps γ_1, γ_2, and identity maps, as shown:

$$\begin{array}{ccccc}
H_*^1(L, M) & \to & H^1(L, M) & \to & H^1(C(M)/C^0(M)) \\
\updownarrow & & \updownarrow & & \downarrow \gamma_1 \\
H_*^1(L, M) & \to & H^1(L, M) & \to & S(L, M^L)
\end{array}$$

$$\begin{array}{ccccccc}
\to H_*^2(L, M) & \to & H^2(L, M) & \to & H^2(C(M)/C^0(M)) \\
\updownarrow & & \updownarrow & & \downarrow \gamma_2 \\
\to H_*^2(L, M) & \to & H^2(L, M) & \to & S(L, H^1(L, M)).
\end{array}$$

It can be verified from our definitions of these maps that this diagram

is commutative, and a familiar argument shows that γ_1 is an isomorphism onto and that γ_2 maps the image of $H^2(L, M)$ in $H^2(C(M)/C^0(M))$ isomorphically onto the image of $H^2(L, M)$ in $S(L, H^1(L, M))$.

This suggests that the bottom sequence may be continued, preserving exactness, with a map $S(L, H^1(L, M)) \to H_*{}^3(L, M)$. Such a continuation would be of interest in connection with the interpretations of the 3-dimensional cohomology groups with which we intend to deal in a later paper, but to which the next section is already an introduction.

Finally, we observe that it should be possible to prove these cohomology results, without making any use of the interpretations of the cohomology groups, by filtering the associative cochain complex for $R_L{}^+$ in M relative to the ideal P_L and considering the spectral sequence which arises from such a filtration. This procedure would amount to a slight modification of that of Kung-sing Shih, in [7].

4. Non-abelian kernels. In the case of groups, it is well known that the enumeration of the extensions of an arbitrary kernel can be reduced to that of the extensions of the center of that kernel.[10] The analogue of this reduction, for the case of associative algebras, is contained in [4], and it is clear that the case of Lie algebras can be dealt with in the same way. In the case of restricted Lie algebras, one has to augment the ordinary procedure by the appropriate considerations relating to the p-maps. We shall briefly indicate the resulting theory for this case. The case of ordinary Lie algebras will then be contained implicitly in our discussion; one merely omits everything which relates to the p-maps.

We begin by recalling some of Jacobson's results on derivations of restricted Lie algebras. A derivation ζ of the restricted Lie algebra M is said to be a restricted derivation if $\zeta(m^{[p]}) = D_m{}^{p-1}(\zeta(m))$, for every $m \in M$. Observe that every derivation ζ of M can be extended uniquely to a derivation ζ^* of R_M which annihilates the base field F. One sees immediately that ζ is a restricted derivation of M if and only if ζ^* annihilates the elements of the form $m^p - m^{[p]}$. From this it is evident that if ζ and τ are restricted derivations of M, so are ζ^p and $[\zeta, \tau] = \zeta\tau - \tau\zeta$. Hence the space $D_0(M)$ consisting of the restricted derivations of M, with the usual commutation, and the p-map $\zeta \to \zeta^{[p]} = \zeta^p$, constitute a restricted Lie algebra.

Now let L be another restricted Lie algebra, and consider a restricted extension (E, ϕ) of M by L. Choose a linear map ρ of L into E which is

[10] See [3], which also gives an account of similar results due to R. Baer.

inverse to ϕ. With each $x \in L$, associate the derivation ζ of M which is effected by the element $\rho(x)$ of E, i. e., $\zeta(m) = [\rho(x), m]$. Condition (1) on the p-map in E implies that ζ is then a restricted derivation of M. Let $I(M)$ denote the set of the inner derivations of M. It is immediate that $I(M)$ is a restricted ideal of $D_0(M)$, in the sense that $I(M)^{[p]} \subset I(M)$. The above correspondence $x \to \zeta$ induces a homomorphism of L into $D_0(M)/I(M)$. Furthermore, it is easy to verify that this homomorphism is a restricted homomorphism of L into the restricted Lie algebra $D_0(M)/I(M)$. Finally, this homomorphism evidently depends only on the equivalence class of the restricted extension (E, ϕ).

We suppose that we are given a restricted Lie algebra M and a restricted homomorphism of L into the restricted Lie algebra $D_0(M)/I(M)$. We shall refer to this structure by saying that we are given a restricted L-kernel M. If N denotes the center of M, then N is a restricted abelian Lie algebra on which L operates, in the sense defined in the previous section. Clearly, it means the same if we refer to this structure by saying that we have a restricted abelian L-kernel. We shall call this the nucleus of the restricted L-kernel M.

In accordance with our earlier notation, we shall denote by $\mathrm{ext}_*(M, L)$ the set of equivalence classes of those restricted extensions of M by L which induce the given L-kernel structure on M. The chief tool for our reduction of $\mathrm{ext}_*(M, L)$ to $\mathrm{ext}_*(N, L)$, where N is the nucleus of M, is a composition which associates with an element $u \in \mathrm{ext}_*(M, L)$, and an element $v \in \mathrm{ext}_*(N, L)$, an element $u + v \in \mathrm{ext}_*(M, L)$. As the notation suggests, this is an evident extension of the addition in $\mathrm{ext}_*(N, L)$. We shall describe this composition briefly.

Let (E, ϕ) be a restricted extension of M by L which belongs to the equivalence class u, and let (T, τ) be a restricted extension of N by L which belongs to v. Let D denote the subalgebra of the direct sum of E and T which consists of all elements (e, t) in which $\phi(e) = \tau(t)$. Let J be the ideal of D consisting of the elements $(n, -n)$, where $n \in N$. In D/J, we identify $(M, N)/J$ with M by using the homomorphism $(m, n) \to m + n$, whose kernel is exactly J. The map $(e, t) \to \phi(e)$ induces a homomorphism ψ of D/J onto L whose kernel is M. The map $(e, t) \to (e^{[p]}, t^{[p]})$ induces a p-map in D/J, because it sends J into itself. With this p-map, $(D/J, \psi)$ constitutes a restricted extension of M by L whose equivalence class is easily seen to depend only on u and v, and which is the desired composite $u + v$. It is easy to verify that this composition has the properties suggested by the

notation. In particular, $u + 0 = u$, and $(u + v) + v' = u + (v + v')$. Also, if $u + v = u$, it follows that $v = 0$.

The reduction theorem can now be expressed as follows:

THEOREM 4.1. *Let L be a restricted Lie algebra and let M be a restricted L-kernel. Let u be any element of $\mathrm{ext}_*(M, L)$, and let N denote the nucleus of M. Then the map $v \to u + v$ of $\mathrm{ext}_*(N, L)$ into $\mathrm{ext}_*(M, L)$ is* 1-1 *and sends $\mathrm{ext}_*(N, L)$ onto $\mathrm{ext}_*(M, L)$.*

Proof. We shall prove only the second statement in detail. Let (E, ϕ) be a restricted extension of M by L which belongs to the given equivalence class u. Let E^* be the restricted Lie algebra which is obtained from E simply by changing the sign in the commutation. More precisely, we have an anti-isomorphism $e \to e^*$ such that $[e, f]^* = -[e^*, f^*]$ and $(e^*)^{[p]} = (e^{[p]})^*$. Let (E_1, ϕ_1) be any restricted extension of M by L which induces the given L-kernel structure in M. Let D denote the subalgebra of the direct sum of E_1 and E^* which consists of those elements $(-e_1, e^*)$ in which $\phi_1(e_1) = \phi(e)$ and also $[e_1, m] = [e, m]$, for all $m \in M$. D is a restricted Lie algebra (with the p-map defined component-wise, as before) and the map $(-e_1, e^*) \to -\phi(e)$ is a restricted homomorphism of D into L. Using that (E_1, ϕ_1) and (E, ϕ) induce the same L-kernel structure on M, it is easy to see that this homomorphism actually maps D onto L. The kernel K of this homomorphism consists of all elements of the form $(-m, (m + n)^*)$, where $m \in M$ and $n \in N$. Let J be the ideal of D consisting of the elements $(-m, m^*)$, with $m \in M$. In D/J, we identify K/J with N by means of the homomorphism $(-m, (m + n)^*) \to n$ whose kernel is precisely J. Since the p-map in D sends J into itself, D/J is a restricted Lie algebra. If τ is the restricted homomorphism of D/J onto L which is induced by the above map $(-e_1, e^*) \to -\phi(e)$, then $(D/J, \tau)$ is a restricted extension of N by L which represents an element $v \in \mathrm{ext}_*(N, L)$.

We claim that the equivalence class of (E_1, ϕ_1) is the composite $u + v$. In order to see this, we must exhibit an equivalence isomorphism of the composite (G, γ), say, of (E, ϕ) with $(D/J, \tau)$ onto (E_1, ϕ_1). An element $g \in G$ is represented by a triple $(e', -e_1, e^*)$, where $\phi_1(e_1) = \phi(e) = -\phi(e')$. Hence $e + e' \in M$, and we can form $(e + e') - e_1$ in E_1. It is easily seen that this element of E_1 does not depend on the particular choice of the representative triple, but only on g. Hence we may define $\alpha(g) = (e + e') - e_1$. Clearly, we have $\phi_1 \alpha = \gamma$. An element $m \in M$ is represented by the triple $(m, 0, 0)$, whence it is clear that α leaves the elements of M fixed. Evidently,

α maps G linearly onto E_1. If $\alpha(g) = 0$ we have $e_1 = (e + e') \in M$, whence also $e \in M$ and $e' \in M$. Hence $g \in M$, and therefore $g = \alpha(g) = 0$. Hence α is 1-1. Now let g_1 and g_2 be two arbitrary elements of G and choose representative triples $(e', -e_1, e^*)$ and $(f', -f_1, f^*)$ for g_1 and g_2, respectively. Then $[g_1, g_2]$ is represented by the triple $([e', f'], [e_1, f_1], -[e, f]^*)$. Hence $\alpha([g_1, g_2]) = ([e', f'] - [e, f]) + [e_1, f_1]$. On the other hand,

$$[\alpha(g_1), \alpha(g_2)] = [(e + e') - e_1, (f + f') - f_1]$$
$$= [(e + e') - e_1, f + f'] - [(e + e') - e_1, f_1].$$

Using that e and e_1, and f and f_1, effect the same derivations in M, this can be written

$$([e', f + f'] - [(e + e'), f]) + [e_1, f_1]$$
$$= ([e', f'] - [e, f]) + [e_1, f_1] = \alpha([g_1, g_2]).$$

Hence α is an isomorphism of G onto E_1. Finally, we have

$$(\alpha(g))^{[p]} = (e + e')^{[p]} - e_1^{[p]} + \sigma(e + e', -e_1),$$

where σ is the map which occurs in condition (3) on p-maps. We know that $\sigma(e + e', -e_1)$ is the formal integral of the polynomial

$$D_{(e+e')t - e_1}{}^{p-1}(e + e').$$

Since e and e_1 effect the same derivation in M, we may here replace e_1 by e. It follows that

$$\sigma(e + e', -e_1) = \sigma(e + e', -e) = (e')^{[p]} - (e + e')^{[p]} + e^{[p]}.$$

Hence $(\alpha(g))^{[p]} = ((e')^{[p]} + e^{[p]}) - e_1^{[p]} = \alpha(g^{[p]})$. This shows that α is an equivalence isomorphism, and Theorem 4. 1 is proved.

As in the case of groups, [3], and associative algebras, [4], the study of L-kernels M leads to 3-dimensional cohomology. The procedure analogous to that of [3] and [4] leads to 3-cocycles of the Lie type for L in the nucleus N of M. In order to extend the technique used above, it would be necessary to establish a direct connection between L-kernels and 3-cocycles of the associative type. There are a number of unsolved problems in this direction.

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] C. Chevalley, *Théorie des groupes de Lie*, vol. II, Paris, 1951.

[2] B. Eckmann, "On complexes over a ring and restricted cohomology groups," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 275-281.

[3] S. Eilenberg and S. MacLane, "Cohomology theory in abstract groups, II," *Annals of Mathematics*, vol. 48 (1947), pp. 326-341.

[4] G. Hochschild, "Cohomology and representations of associative algebras," *Duke Mathematical Journal*, vol. 14 (1947), pp. 921-948.

[5] G. Hochschild and J-P. Serre, "Cohomology of Lie algebras," *Annals of Mathematics*, vol. 57 (1953), pp. 591-603.

[6] N. Jacobson, "Restricted Lie algebras of characteristic p," *Transactions of the American Mathematical Society*, vol. 50 (1941), pp. 15-25.

[7] Kung-sing Shih, *Cohomology of associative algebras and spectral sequences*, Thesis, University of Illinois, 1953.

NORMAL COMPLETIONS OF INCIDENCE MATRICES.*

By MARSHALL HALL and H. J. RYSER.

---

1. Introduction. In this paper we consider the problem of arranging v elements into v sets, such that each set contains exactly k distinct elements and such that every pair of sets has exactly $\lambda = k(k-1)/(v-1)$ elements in common $(0 < \lambda < k < v)$. The significance of the v, k, λ configuration is described in detail in [3]. In particular for $v = N^2 + N + 1$, $k = N + 1$, $N \geqq 2$, and $\lambda = 1$, the problem is equivalent to determining a projective plane with $N + 1$ points on a line. For $v = 4N - 1$, $k = 2N - 1$, and $\lambda = N - 1$, the problem reduces to a determination of a Hadamard matrix of order $4N$. Finally, we remark that the general v, k, λ configuration arises in a natural way in statistics, where the arrangements are referred to as symmetrical balanced incomplete block designs.

Let the elements $x_1, \cdots, x_v$ be listed in a row and let the sets $s_1, \cdots, s_v$ be listed in a column. We form the incidence matrix A of the configuration in the usual way, by inserting 1 in row i and column j if x_j belongs to s_i and 0 otherwise. The matrix A is of order v and is composed entirely of 0's and 1's. Clearly A satisfies the matric equation $AA^T = B$. Here A^T denotes the transpose of A, and B is the matrix with k in the main diagonal and λ in all other positions. For incidence matrices, one readily proves that $A^TA = B$, so that the incidence matrix is normal.

It is evident that the matrix B is congruent to the identity, relative to the field of rational numbers. Bruck and Ryser have utilized this fact for the establishment of the theorem on the nonexistence of finite projective planes [2]. Later this result was extended by Chowla and Ryser to the general v, k, λ configuration, and the proof of the nonexistence theorem was made independent of the Minkowski-Hasse theory of the invariants of a quadratic form [3]. Shrikhande, basing his investigations directly on [2], independently derived the result on the nonexistence of v, k, λ configurations [8]. Further results, with emphasis on the cyclic case, may be found in Hall and Ryser [4]. In summary, a necessary and sufficient condition in order that the matrix B of even order v be congruent to the identity is that $k - \lambda$ be a square. For v odd, the condition is that the norm residue symbols

* Received November 19, 1953.

satisfy $(k-\lambda,-1)_p^{v(v-1)/2}\ (k-\lambda,v)_p=+1$ for every odd prime p. The latter condition is equivalent to asserting that the Diophantine equation $x^2=(k-\lambda)y^2+(-1)^{(v-1)/2}\lambda z^2$ possesses a nonzero solution in integers.

For the case of the non-symmetric block design, however, the condition for rational solvability is usually vacuous. But Connor [5] has given certain conditions which must be satisfied if a given initial set of blocks can be completed to a design. Connor's conditions may be interpreted as the requirement that a certain quadratic form may not be indefinite. Connor's conditions do not yield results for symmetric designs.

In an effort to obtain results similar to Connor's for symmetric designs, the authors of this paper found entirely contrary results, which are given here. The rational conditions for solvability being satisfied, there exists (a) a normal and rational matrix satisfying the incidence equation and (b) any set of initial rows satisfying the trivially necessary conditions can be rationally completed to a normal solution of the incidence equation. Of these results, the first generalizes a theorem of Albert [1] and the second is in strong contrast to the results of Connor.

2. Matrix preliminaries. Let A, B, C be symmetric matrices with elements in a field F of characteristic $\neq 2$. Suppose that $C\oplus A\underline{\subset}C\oplus B$, where the notation $\oplus$ denotes direct sum and $\underline{\subset}$ denotes congruence in the sense that there exists a nonsingular X with elements in F such that $X^{\mathsf{T}}(C\oplus A)X=C\oplus B$. Then Witt's theorem [9] asserts that $A\underline{\subset}B$. Theorem 2.1 which follows is closely related to a theorem of G. Pall [6], and will be derived here by the use of Witt's theorem.

THEOREM 2.1. *Suppose that $AA^{\mathsf{T}}=D_1\oplus D_2$. Here the matrix A is of order n and nonsingular. The matrix D_1 is of order r and D_2 is of order s, where $r+s=n$. Let X be an arbitrary $r\times n$ matrix such that $XX^{\mathsf{T}}=D_1$. Then there exists an $n\times n$ matrix Z having X as its first r rows such that $ZZ^{\mathsf{T}}=D_1\oplus D_2$. This result holds for all fields F of characteristic $\neq 2$.*

The matrix X is of rank r since D_1 is of rank r. Thus we may take an $s\times n$ matrix W of rank s whose rows are a basis for the space annihilated by X. Then letting

$$Y=\begin{bmatrix}X\\W\end{bmatrix},$$

we have $YY^{\mathsf{T}}=D_1\oplus D_3$. We assert that Y is nonsingular. There exists

a nonsingular C such that $CD_1C^T = D$, where $D = \text{diag}[d_1, \cdots, d_r]$, with $d_i \neq 0$. Thus

$$\begin{bmatrix} CX \\ W \end{bmatrix} [X^TC^T, W^T] = \begin{bmatrix} D & 0 \\ 0 & D_3 \end{bmatrix}.$$

Let the row vectors of

$$T = \begin{bmatrix} CX \\ W \end{bmatrix}$$

be denoted by $\alpha_1, \cdots, \alpha_r, \beta_1, \cdots, \beta_s$. We assert that T is nonsingular. For if $a_1\alpha_1 + \cdots + a_r\alpha_r + b_1\beta_1 + \cdots + b_s\beta_s = 0$, then since the inner product $(\alpha_i, \alpha_j) = d_i\delta_{ij}$, with $d_i \neq 0$ and δ_{ij} the Kronecker delta, it follows that $a_1 = \cdots = a_r = 0$. But since the β's are linearly independent, we have $b_1 = \cdots = b_s = 0$. Thus the determinant $| D_3 | \neq 0$ so that Y is nonsingular. By Witt's theorem, there is a nonsingular E such that $ED_3E^T = D_2$. But then with

$$Z = \begin{bmatrix} X \\ EW \end{bmatrix},$$

we have

$$ZZ^T = \begin{bmatrix} D_1 & 0 \\ 0 & ED_3E^T \end{bmatrix} = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}.$$

THEOREM 2.2. *Let $x = (x_1, \cdots, x_n)$ and let $y = (y_1, \cdots, y_n)$ be vectors such that*

$$x_1^2 + \cdots + x_n^2 = y_1^2 + \cdots + y_n^2 = c \neq 0.$$

Then there exists an orthogonal matrix O such that $xO = y$. This result holds for all fields F of characteristic $\neq 2$.

There exists an $(n-1) \times n$ matrix W of rank $n-1$ such that

$$X = \begin{bmatrix} x \\ W \end{bmatrix} \text{ satisfies } XX^T = \begin{bmatrix} c & 0 \\ 0 & D_2 \end{bmatrix}.$$

Since $c \neq 0$, it follows as in the proof of Theorem 2.1, that X is nonsingular. By Theorem 2.1, there then exists a matrix

$$Y = \begin{bmatrix} y \\ R \end{bmatrix} \text{ such that } YY^T = \begin{bmatrix} c & 0 \\ 0 & D_2 \end{bmatrix}.$$

Evidently the matrix $O = X^{-1}Y$ is orthogonal and we may write

$$(1, 0, \cdots, 0)XO = (1, 0, \cdots, 0)Y.$$

Thus it follows that $xO = y$.

Theorem 2. 2 has to do with the preservation of length by rotations. This is well known for the field of reals, but, in particular for the rational field to which we wish to apply this result, it is not obvious (without Theorem 2. 2) that we can find a rational rotation carrying one vector into another of the same length.

3. Incidence matrices. We return now to the topic of incidence matrices, and in the following, B always denotes the $v \times v$ matrix with k in the main diagonal and $\lambda = k(k-1)/(v-1)$ in all other positions.

THEOREM 3. 1. *Suppose that there is a rational C such that $CC^T = B$. Suppose, moreover, that $CS = kS$, where S is the matrix of all ones. Then $CC^T = B$.*

For by hypothesis,

$$C^TC = B = (k-\lambda)I + \lambda S \tag{3.1}$$

so that

$$C^TCS = (k-\lambda)S + \lambda vS. \tag{3.2}$$

Thus

$$C^TS = [(k-\lambda) + \lambda v]k^{-1}S = kS \tag{3.3}$$

and, transposing, we have

$$SC = kS. \tag{3.4}$$

Hence also

$$SC^{-1} = k^{-1}S. \tag{3.5}$$

From (3. 1) and (3. 5), we have

$$(k-\lambda)C^{-1} = C^T - \lambda k^{-1}S \tag{3.6}$$

and multiplying (3. 6) on the left by C gives

$$(k-\lambda)I = CC^T - \lambda k^{-1}CS, \tag{3.7}$$

whence

$$CC^T = B. \tag{3.8}$$

The following is a generalization of a theorem of Albert [1].

THEOREM 3. 2. *Suppose that the matrix B is congruent to the identity, relative to the field of rational numbers. Then there exists a rational C such that $CC^T = C^TC = B$.*

By hypothesis, there exists a rational A such that

$$A^{\mathsf{T}}A = B. \tag{3.9}$$

Let S again denote the matrix of all ones, and let r_i denote the sum of the i-th row of A. Then

$$SA^{\mathsf{T}}AS = SBS = k^2vS \tag{3.10}$$

so that, calculating the element in the $(1, 1)$ position on the right, we have

$$r_1^2 + \cdots + r_v^2 = k^2v. \tag{3.11}$$

Hence by Theorem 2.2, there exists an orthogonal O such that

$$O\begin{bmatrix} r_1 \\ \vdots \\ r_v \end{bmatrix} = \begin{bmatrix} k \\ \vdots \\ k \end{bmatrix}. \tag{3.12}$$

Consider the matrix

$$C = OA. \tag{3.13}$$

Evidently,

$$C^{\mathsf{T}}C = A^{\mathsf{T}}O^{\mathsf{T}}OA = B. \tag{3.14}$$

Moreover, by (3.12), we have

$$CS = O\begin{bmatrix} r_1, \cdots, r_1 \\ \vdots \qquad\quad \vdots \\ r_v, \cdots, r_v \end{bmatrix} = kS. \tag{3.15}$$

By Theorem 3.1, this proves normality.

We come now to our main result.

THEOREM 3.3. *Suppose that the matrix B is rationally congruent to the identity. Let A_1 be an $r \times v$ matrix of 0's and 1's such that $A_1A_1^{\mathsf{T}} = B_1$, where B_1 is the $r \times r$ matrix with k in the main diagonal and λ elsewhere. Then there exists a $v \times v$ rational matrix A having A_1 as its first r rows such that $AA^{\mathsf{T}} = A^{\mathsf{T}}A = B$.*

Concerning Theorem 3.3, we remark that if we are given r rows of a v, k, λ configuration, then a trivial condition necessary for completion is that their pairwise intersections equal λ and their sums equal k. This is the condition

$$A_1A_1^{\mathsf{T}} = B_1 \tag{3.16}$$

of the theorem.

The rational matrix C of Theorem 3.2 satisfies

$$C^{\mathsf{T}}C = CC^{\mathsf{T}} = B \tag{3.17.1}$$

and

$$CS = SC = kS. \tag{3.17.2}$$

Consider the $r \times v$ matrix A_1C^{-1}. This matrix satisfies

$$\begin{aligned}(A_1C^{-1})(A_1C^{-1})^{\mathsf{T}} &= A_1C^{-1}C^{-1\mathsf{T}}A_1^{\mathsf{T}} = A_1B^{-1}A_1^{\mathsf{T}} \\ &= A_1[(k-\lambda)^{-1}(I-\lambda k^{-2}S)]A_1^{\mathsf{T}} = (k-\lambda)^{-1}B_1 - \lambda(k-\lambda)^{-1}S_1,\end{aligned} \tag{3.18}$$

where S_1 is the $r \times r$ matrix of all ones. Here we have used not only (3.16), but also the fact that the rows of A_1 sum to k. Thus

$$(A_1C^{-1})(A_1C^{-1})^{\mathsf{T}} = I, \tag{3.19}$$

where I is the $r \times$ r identity. Hence by Theorem 2.1, there exists a rational Q such that

$$\begin{bmatrix} A_1C^{-1} \\ Q \end{bmatrix} \tag{3.20}$$

is orthogonal. But this implies that

$$[C^{-1\mathsf{T}}A_1^{\mathsf{T}}, Q^{\mathsf{T}}]\begin{bmatrix} A_1C^{-1} \\ Q \end{bmatrix} = I \tag{3.21}$$

or

$$[A_1^{\mathsf{T}}, C^{\mathsf{T}}Q^{\mathsf{T}}]\begin{bmatrix} A_1 \\ QC \end{bmatrix} = C^{\mathsf{T}}C = B. \tag{3.22}$$

Let r_i denote the sum of the i-th row of QC. We assert that

$$r_1^2 + \cdots + r_s^2 = k^2 s, \tag{3.23}$$

where $s = v - r$. For we have

$$A_1^{\mathsf{T}}A_1 + (QC)^{\mathsf{T}}QC = B, \tag{3.24}$$

whence

$$SA_1^{\mathsf{T}}A_1S + S(QC)^{\mathsf{T}}QCS = k^2vS, \tag{3.25}$$

and as A_1 has r rows summing to k,

$$S(QC)^{\mathsf{T}}(QC)S = k^2(v-r)S, \tag{3.26}$$

proving (3.23).

By Theorem 2.2, there exists an $s \times s$ orthogonal matrix O such that

$$O\begin{bmatrix} r_1 \\ \vdots \\ r_s \end{bmatrix} = k\epsilon_1, \tag{3.27}$$

where ϵ_1 is the column vector of s components all equal to one. Moreover,

$$(QC)\epsilon = \begin{bmatrix} r_1 \\ \vdots \\ r_s \end{bmatrix}, \tag{3.28}$$

where ϵ is the column vector of v components all equal to one. Hence it follows that

$$(OQC)\epsilon = k\epsilon_1. \tag{3.29}$$

We now define

$$A = \begin{bmatrix} A_1 \\ OQC \end{bmatrix}. \tag{3.30}$$

Evidently,

$$A^{\mathsf{T}}A = B. \tag{3.31}$$

But also

$$A\epsilon = k\epsilon, \tag{3.32}$$

and the normality of A follows by Theorem 3.1.

4. Integral solutions. In this section we consider the solutions of the matric equation $AA^{\mathsf{T}} = B$, where A is integral. We remark in passing that it is easy to show that if also $A^{\mathsf{T}}A = B$, then A is composed entirely of 0's and 1's or entirely of 0's and -1's [7].

Suppose now that $AA^{\mathsf{T}} = B$, where A is integral. In this section we suppose further that k and $k - \lambda$ are always relatively prime. Then if $k - \lambda$ is odd, it has been shown that upon suitable multiplication of the columns of A by ± 1, A is transformed into an incidence matrix composed entirely of 0's and 1's [7]. The case of $k - \lambda$ even, however, presents additional difficulties. We confine ourselves for the most part to the case of the projective plane, with $k - \lambda = N$ and $v = N^2 + N + 1$.

Let s_i denote the sum of the i-th column of A and let t_{ij} denote the inner product of the i-th and j-th columns of A. We agree to write A so that we always have $s_i \geqq 0$. In [7] it was shown that $s_i \equiv 0 \pmod{k}$ and if we write $s_i = u_i k$, it was also shown that $u_1^2 + \cdots + u_v^2 = v$ and $t_{ij} = \lambda u_i u_j + (k - \lambda)\delta_{ij}$, where δ_{ij} is the Kronecker delta.

For the projective plane case, the integral solutions of $AA^{\mathsf{T}} = B$ are of two types. A solution of type *I* is defined by the conditions that $u_i = 1$ for each i. A solution of type *II* is defined by the condition that $u_i \neq 1$ for some i. For type *I*, we have $s_i = N + 1$ and $t_{ii} = N + 1$ for all i. It follows easily that solutions of type *I* are incidence matrices.

For solutions of type *II*, we cannot have $u_i > N$ because $u_1^2 + \cdots + u_v^2 = N^2 + N + 1$. Since $s_i \leqq t_{ii} = u_i^2 + N$ and $u_i^2 - (N+1)u_i + N \geqq 0$, we must have some u_i, say u_1, such that $u_1 = N$. But then there are $N+1$ u_i such that $u_i = 1$ and $N^2 - 1$ u_i such that $u_i = 0$. Then $s_1 = N^2 + N$ and $t_{11} = N^2 + N$. The first column of A thus has a single 0, and all other entries $+1$. For convenience, the 0 may be placed in the $(1,1)$ position of A. In addition, there are $N+1$ columns with $s_i = t_{ii} = N+1$. These columns have $N+1$ ones, and all other entries 0. For these columns $t_{1i} = N \cdot 1 = N$, whence one of the 1's is in the first row. Thus the first row consists of $N+1$ 1's and other entries 0.

Upon deleting the first row and column of A, we obtain a matrix C of order $N^2 + N$ such that $CC^\intercal = C^\intercal C = NI$. There are $N+1$ columns of C composed entirely of 0's and 1's. Conversely, if these reguirements are fulfilled, then it is easy to construct the integral solution A of type *II*. By the results of § 2, it is clear that it is always possible to find rational matrices fulfilling these requirements, provided, of course, that NI is rationally congruent to the identity.

It is much more difficult, however, to determine those values of N for which there are integral solutions of type *II*. We conjecture that such solutions exist whenever N is even, provided only that B is rationally congruent to the identity. To substantiate this conjecture, integral solutions are known to exist whenever N is the order of a Hadamard matrix [7]. This, however, restricts N to the value of 2 or to $N \equiv 0 \pmod 4$. In conclusion, we exhibit an integral solution for $N = 10$. Let

$$X = \begin{pmatrix}
1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & -2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 2 & -1 & -1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & -1 & 2 & -1 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & -1 & -1 & 2 & 0 & 0 & 0 & 1 \\
1 & -1 & 1 & 0 & 0 & 0 & 2 & -1 & -1 & -1 \\
1 & -1 & 0 & 1 & 0 & 0 & -1 & -2 & 1 & 1 \\
1 & -1 & 0 & 0 & 1 & 0 & -1 & 1 & -2 & 1 \\
1 & -1 & 0 & 0 & 0 & 1 & -1 & 1 & 1 & -2
\end{pmatrix}.$$

The matrix X is such that $XX^\intercal = 10I$. The matrix C may be formed from X by taking a direct sum of X with itself eleven times. From C the desired A of order 111 is obtained by adjoining the suitable row and column.

THE OHIO STATE UNIVERSITY.

BIBLIOGRAPHY.

[1] A. A. Albert, "Rational normal matrices satisfying the incidence equation," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 554-559.

[2] R. H. Bruck and H. J. Ryser, "The nonexistence of certain finite projective planes," *Canadian Journal of Mathematics*, vol. 1 (1949), pp. 88-93.

[3] S. Chowla and H. J. Ryser, "Combinatorial problems," *Canadian Journal of Mathematics*, vol. 2 (1950), pp. 93-99.

[4] Marshall Hall and H. J. Ryser, "Cyclic incidence matrices," *Canadian Journal of Mathematics*, vol. 3 (1951), pp. 495-502.

[5] W. S. Connor, Jr., "On the structure of balanced incomplete block designs," *Annals of Mathematical Statistics*, vol. 23 (1952), pp. 57-71.

[6] Gordon Pall, "Hermitian quadratic forms in a quasi-field," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 889-893.

[7] H. J. Ryser, "Matrices with integer elements in combinatorial investigations," *American Journal of Mathematics*, vol. 74 (1952), pp. 769-773.

[8] S. S. Shrikhande, "The impossibility of certain symmetrical balanced incomplete block designs," *Annals of Mathematical Statistics*, vol. 21 (1950), pp. 106-111.

[9] Ernst Witt, "Theorie der quadratischen Formen in beliebigen Körpern," *Journal für die reine und angewandte Mathematik*, vol. 176 (1937), pp. 31-44.

ON THE λ-DIMENSION AND THE *A*-DIMENSION OF PARTIALLY ORDERED SETS.*

By SEYMOUR GINSBURG.

Let S be an abstract set and K a collection of linear orders defined on S. Define a partial order P on S as follows. For any two elements p and q in S put $p < q$ in P if and only if $p < q$ for each linear order in K. The partial order P so obtained is said to be realized by the linear orders in K. By the dimension of a partial order P on a set S is meant the smallest cardinal number m such that P is realized by m linear orders.

The above definition of 'dimension of a partially ordered set' was first stated by Dushnik and Miller in [2]. They showed that each partial order has a dimension. Other papers pertaining to this definition of dimension are [1], [3], [6], and [8]. In [6] Komm defined the λ-dimension of a partial order P on a set S as the minimum number of linear extension of P[1] each of which maps S into the real numbers. He then calculated the λ-dimension of several partial orders. The purpose of this note is twofold:

(1) To make some remarks on λ-dimension. In particular, the λ-dimension of the partially ordered set consisting of sequences of real numbers, ordered as follows, is shown to be $\aleph_0$: $(x_1, \cdots, x_n, \cdots) \leqq (y_1, \cdots, y_n, \cdots)$ if and only if $x_n \leqq y_n$ for all n.

(2) To discuss a new kind of dimension, called *A-dimension*. A non-trivial example is given to show that λ-dimension and A-dimension need not be equal, even when both dimensions exist.

1. λ-dimension. For each positive integer n denote by E_n the set of n-tuples $E_n = \{(x_1, \cdots, x_i, \cdots, x_n) \mid x_i \text{ a real number}\}$. Denote by E_∞ the set $E_\infty = \{(x_1, \cdots, x_i, \cdots) \mid x_i \text{ a real number}\}$. Let $P(E_n)$ be the partial order on E_n which is defined as follows: $(x_1, \cdots, x_n) < (y_1, \cdots, y_n)$ if and only if $x_i < y_i$ for each i. Let $P(E_\infty)$ be the partial order on E_∞ which is

* Received July 24, 1953.

[1] The linear order L on a set S is a linear extension of a partial order P if, for any two elements p and q such that $p<q$ in P, then $p<q$ in L. If L is similar to a subset of the real numbers, then L is called a λ-extension of P.

defined as follows: $(x_1, \cdots, x_i, \cdots) < (y_1, \cdots, y_i, \cdots)$ if and only if $x_i < y_i$ for each i. Let $P'(E_n)$ and $P'(E_\infty)$ be the partial orders on E_n and E_∞ respectively which are defined as $P(E_n)$ and $P(E_\infty)$ except that "$x_i < y_i$ for each i" is replaced by "$x_i \leqq y_i$ for each i, and for some $j, x_j < y_j$." Let M be a subset of E_n or E_∞. By $P(M)$ is meant the partial order $P(E_n)$ or $P(E_\infty)$ acting on M, and by $P'(M)$ is meant $P'(E_n)$ or $P'(E_\infty)$ on M, according as M is a subset of E_n or E_∞.

In [6] Komm demonstrated that the λ-dimension of $P'(E_n)$ is $\aleph_0$. Komm's method however cannot be extended to calculate the λ-dimension of $P'(E_\infty)$. In Theorem 1.1 we shall show, by a procedure which is also capable of calculating the λ-dimension of $P'(E_n)$, that the λ-dimension of $P'(E_\infty)$ is $\aleph_0$.

Denote by J the open interval $(0, 1)$ of the real numbers. Denote by J_∞ the set $J_\infty = \{(x_1, \cdots, x_n, \cdots) \mid \text{each } x_n \in J\}$.

LEMMA 1.1. *There exists a mapping ψ of J into $P'(E_\infty)$ such that*

(α) $x^{(i)} \leqq 2^{i+1}$, *where* $\psi(x) = (x^{(1)}x^{(2)}, \cdots, x^{(n)}, \cdots)$, *the* $x^{(i)}$ *being positive integers,*

(β) *if* $x < y$ *in the real number system, then* $\psi(x) < \psi(y)$ *in* $P'(E_\infty)$, *and*

(γ) ψ *is one to one.*

Proof. Each element of J can be written uniquely as $x = \sum_{i=1}^{\infty} u_i/2^i$, where each $u_i = 0$ or 1, and the u_i are not ultimately all 1. Let $x^{(1)} = u_1 + 1$ and, in general, let $x^{(n+1)} = 2x^{(n)} + u_{n+1} + 1$. Let $\psi(x) = (x^{(1)}, \cdots, x^{(n)}, \cdots)$. There is no difficulty in verifying that the function ψ satisfies the conclusions of the lemma.

THEOREM 1.1. *The λ-dimension of $P'(E_\infty)$ is $\aleph_0$.*

Proof. Since J is similar to the real numbers it follows that the two partially ordered sets, $P'(J_\infty)$ and $P'(E_\infty)$ are isomorphic. Now two isomorphic partially ordered sets evidently have the same λ-dimension. Consequently, in order to demonstrate the theorem, it is sufficient to show that the λ-dimension of $P'(J_\infty)$ is $\aleph_0$.

(a) We shall now construct the auxiliary functions, f, g, h, and for i and j any positive integers such that $i < j$, the function p_{ij}. In (b) these functions will play an important role in calculating the λ-dimension of $P'(J_\infty)$.

To define f we first decompose the set of positive integers into a family

$\{W_n/1 \leqq n < \omega\}$ of pairwise disjoint sets, each set W_n being infinite. For each n denote by f_n the similarity transformation which maps the positive integers onto the set W_n, where W_n is ordered according to the natural order of the integers in it. To each element $x = (x_1, \cdots, x_n, \cdots)$ of J_ω we now associate an element $f(x) = (a_1, \cdots, a_n, \cdots)$ of E_ω in the following manner. For each number x_n let $\psi(x_n) = (x^{(1)}{}_n, \cdots, x^{(i)}{}_n, \cdots)$ where ψ is the function defined in Lemma 1. 1. Define $a_{f_n(i)}$ to be the element $x^{(i)}{}_n$. The sets W_n and the functions f_n were constructed in such a manner that for each positive integer k, there exists one and only one pair of integers (n, i) for which $k = f_n(i)$. Therefore the element $f(x) = (a_1, \cdots, a_n, \cdots)$ is well defined. From (α), (β), and (γ) of Lemma 1. 1, and from the fact that each f_n is a similarity transformation onto W_n, we see that the mapping f possesses the following four properties:

- (α) For each x in J_ω, $a_n \leqq 2^{n+1}$, where $f(x) = (a_1, \cdots, a_n, \cdots)$.
- (β) If $x < y$ in $P'(J_\omega)$, then $f(x) < f(y)$ in $P'(E_\omega)$.
- (γ) f is a one to one function.
- (δ) $P'(J_\omega)$ and $P'(f(J_\omega))$ are isomorphic partially ordered sets under the correspondence which takes x into $f(x)$.

We are now going to associate with each element $y = (a_1, \cdots, a_n, \cdots)$ of the set $f(J_\omega)$ a complex $g(y) = (z_1, \cdots, z_n, \cdots)$ of 0's and 4's. Each z_i shall be referred to as a coordinate. The first four coordinates of $g(y)$ shall be denoted as the first set of coordinates. For $i > 1$, the $\sum_{1 \leqq j < i} 2^{j+1} + 1$-th to the $\sum_{1 \leqq j \leqq i} 2^{j+1}$-th, inclusive, coordinates of $g(y)$ shall be denoted as the i-th set of coordinates. The coordinates of the i-th set of coordinates, $i \geqq 1$, are defined as follows. Each of the first a_i of the coordinates is to be a 4. Each of the remaining coordinates is to be a 0. Note that there are 2^{i+1} coordinates in the i-th set of coordinates, and that $a_i \leqq 2^{i+1}$. The function g has the following properties:

- (β') If $x < y$ in $P'(J_\omega)$, then $gf(x) < gf(y)$ in $P'(E_\omega)$.
- (γ') g is a one to one function.
- (δ') $P'(f(J_\omega))$ and $P'(gf(J_\omega))$ are isomorphic partially ordered sets under the correspondence which takes $f(x)$ into $gf(x)$.

We now define the mapping h. For each complex $z = (z_1, \cdots, z_n, \cdots)$ of 0's and 4's let $h(z)$ be the real number $h(z) = \sum_{n=1}^{\infty} z_n/10^n$. The mapping h has the following properties:

(β'') If $z < z'$ in $P'(E_\infty)$, then $h(z) < h(z')$ in the real number system. Furthermore, if $z = (z_1, \cdots, z_n, \cdots)$, $z' = (z'_1, \cdots, z'_n, \cdots)$, and $z \neq z'$, then if i is the first integer for which $z_i \neq z'_i$, and $z_i < z'_i$, then $h(z) < h(z')$.

(γ'') h is a one to one function.

Now let i and j, $i < j$, be any two positive integers. To each complex $a = (a_1, \cdots, a_s, \cdots)$ of 0's and 4's, let $p_{ij}(a) = b = (b_1, \cdots, b_n, \cdots)$, b_n being defined as follows. Let s be an index for a coordinate which is in the m-th set of coordinates of a. If $m < i$ or $m > j$ let $b_s = a_s$. The remaining coordinates of b are to consist of a permutation of the i-th through the j-th set of coordinates of a, the coordinates in each set of coordinates retaining their order in that set of coordinates, under the permutation. The permutation is defined in the following manner. If $a = S_1 S_2 \cdots S_n \cdots$, where S_n is the n-th set of coordinates of a, then

$$b = S_1 S_2 \cdots S_{i-1} S_j S_{i+1} S_{i+2} \cdots S_{j-1} S_i S_{j+1} S_{j+2} \cdots .$$

(b) We now evaluate the λ-dimension of $P'(J_\infty)$. In [6] Komm showed that the λ-dimension of $P'(E_n)$ is $\aleph_0$. Since E_∞ contains a subset which is isomorphic to E_n, the λ-dimension of $P'(J_\infty)$ is at least $\aleph_0$. To complete the proof we shall exhibit $\aleph_0$ λ-extensions of $P'(J_\infty)$ which realize $P'(J_\infty)$.

Consider the function $k = hgf$ which maps J_∞ into the real numbers. In view of (γ), (γ'), and (γ''), k is one to one. Let $x < x'$ in $P'(J_\infty)$. From (β') and (β'') it follows that $k(x) < k(x')$ in the real number system. Consequently the function k is a λ-extension of $P'(J_\infty)$. In a similar manner we see that each of the functions $k_{ij} = hp_{ij}gf$ is a λ-extension of $P'(J_\infty)$. In order to show that the functions k and k_{ij} realize $P'(J_\infty)$, it is sufficient to show that for each pair of incomparable elements x and y in $P'(J_\infty)$, there exist two of the functions, say σ and τ, such that $\sigma(x) < \sigma(y)$ and $\tau(y) < \tau(x)$.

To this end suppose that x and y are two incomparable elements of $P'(J_\infty)$. By (δ) $f(x) = (a_1, \cdots, a_n, \cdots)$ and $f(y) = (b_1, \cdots, b_n, \cdots)$ are two incomparable elements of $P'(E_\infty)$. Let i be the first integer for which $a_i \neq b_i$, say $a_i < b_i$. Since $f(x)$ and $f(y)$ are incomparable elements in $P'(E_\infty)$, there exists an integer j such that $b_j < a_j$. For $s < i$ the s-th set of coordinates of $gf(x)$ is identical with the s-th set of coordinates of $gf(y)$. Since $a_i < b_i$, the first b_i coordinates in the i-th set of coordinates of $gf(y)$ are 4's, whereas the first b_i coordinates in the i-th set of coordinates of $gf(x)$ are a_i 4's and $(b_j - a_i)$ 0's. This, in conjunction with (β'') implies that $k(x) = hgf(x) < hgf(y) = k(y)$. Now consider the two elements $k_{ij}(x)$

and $k_{ij}(y)$. The first a_j coordinates of $p_{ij}gf(x)$ which follow the $\sum_{1\leqq t<i} 2^{t+1}$-th coordinates are all 4's. The first a_j coordinates of $p_{ij}gf(y)$ which follow the $\sum_{1\leqq t<i} 2^{t+1}$-th coordinate are, since $b_j < a_j$, $(a_j - b_j)$ 4's and b_j 0's. For any positive integer $s < \sum_{1\leqq t<i} 2^{t+1}$, if c_s and d_s are the s-th coordinates of $p_{ij}gf(x)$ and $p_{ij}gf(y)$ respectively, then $c_s = d_s$. Consequently

$$k_{ij}(y) = hp_{ij}gf(y) < hp_{ij}gf(x) = k_{ij}(x).$$

Thus the set of λ-extensions of $P'(J_\infty)$, $\{k\} \cup \{k_{ij}/i < j\}$ realize $P'(J_\infty)$. q. e. d.

Remark. The proof of Theorem 1. 1 depended on the construction of $\aleph_0$ λ-extensions of $P'(J_\infty)$. In [7] Ruziewicz defined a function $\phi(x)$ and a sequence of constants $\{c_n\}_{n<\omega}$ such that the mapping σ which takes an element $(x_1, \cdots, x_n, \cdots)$ of J_∞ into the real number $\sum_{i=1}^{\infty} c_i\phi(x_i)$ is a λ-extension of $P'(J_\infty)$. There does not appear to be any way of using Ruziewicz's result to compute the λ-dimension of $P'(J_\infty)$.

Consider the λ-dimension of $P(E_\infty)$. Evidently we may confine ourselves to $P(J_\infty)$. Now if $x < y$ in $P(J_\infty)$, then $x < y$ in $P'(J_\infty)$. This implies that the function k, which was defined in the proof of Theorem 1, 1, is a λ-extension of $P(J_\infty)$. In [6] it was shown that if a partial order P on a set E, of power m, has a λ-extension, then the λ-dimension of P exists and is $\leqq 2m$. Consequently the λ-dimension of $P(J_\infty)$ exists and is $\leqq 2^{\aleph_0}$. Now $P(E_n)$ is of λ-dimension $2^{\aleph_0}$ by 5. 31 of [6]. Since $P(E_\infty)$ contains a set which is isomorphic to $P(E_n)$, theλ-dimension of $P(E_\infty)$ is $\geqq 2^{\aleph_0}$. As a consequence we obtain

THEOREM 1. 2. *The λ-dimension of $P(E_\infty)$ is $2^{\aleph_0}$.*

We conclude this section with some observations in relation to [6]. On p. 517 of [6] there appears the following:

(*) "Let M be any subset of E_n which, except for a denumerable number of verticals [2] has at most one point in common with any vertical. Then the λ-dimension of $P(M)$ and $P'(M)$ is at most n."

For $n > 2$ this result is not true as is now shown. Let M be the set of triplets

$$M = \{(x_1, x_2, x_3) \mid x_1 = 1, x_2 = 1\} \cup \{(x_1, x_2, x_3) \mid x_1 = 3/2, x_2 = 3/2\}.$$

[2] A set V in E_n is a vertical, parallel to the x_i axis, if, for any two elements $x = (x_1, \cdots, x_i, \cdots, x_n)$ and $y = (y_1, \cdots, y_i, \cdots, y_n)$ in V, $x_i \neq y_i$ and $x_j = y_j$ for $j \neq i$.

Clearly M satisfies the hypotheses of (*). For each real number k let $L_k = (1, 1, k)$ and $u_k = (3/2, 3/2, k)$. The arguments given in 6. 2, 6. 3, and 6. 4 of [6] can easily be applied in the present situation to prove that the λ-dimension of $P(M)$ and the λ-dimension of $P'(M)$ are both infinite.

An inspection of the proof of 6. 1 reveals that the argument breaks down at that point when, after projecting M into x_i axis, the claim is made that N_i is similar to a subset of the real numbers. This is not necessarily true. Now 6. 1 can be amended slightly to obtain a correct result. By a vertical plane in E_n, perpendicular to the x_i axis, shall be meant the set of those points $x = (x_1, \cdots, x_n)$ whose i-th coordinate is k, where k is a constant. Then 6. 1 becomes true if the word 'vertical" is replaced by the words 'vertical planes.' In other words

"Let M be a subset of E_n which, except for a denumerable set of vertical planes, has at most one point in common with any vertical plane. Then the λ-dimension of $P'(M)$ and $P(M)$ are both at most n."

The proof given for 6. 1 then suffices with the following changes. Let

$$G_{ik} = \{x \mid x = (x_1, \cdots, x_{i-1}, k, x_{i+1}, \cdots, x_n)\} \cap M.$$

For $P'(M)$, G_{ik} is to be mapped into the interval v associated with the point k by the function f_{ik}, where f_{ik} is a λ-extension of G_{ik} considered in $P'(G_{ik})$. For $P(M)$, G_{ik} is to be mapped into v by the function $g_{ik}f_{ik}$, where g_{ik} is an order reversing mapping of v into itself, i. e., if $x < y$ in v, then $g_{ik}(y) < g_{ik}(x)$ in v.

In connection with the preceding the following generalization of the statement made on lines 20 and 21 of p. 518 of [6] is easily shown.

"Let M be any subset of E_n which has on each of a non-denumerable set of vertical planes at least two comparable points (comparable when considered in the vertical plane). Then the λ-dimension of $P(M)$ and of $P'(M)$ are both infinite."

If 'comparable' is changed to 'incomparable" then the result is no longer true. For example consider the set $M = F \cup G$, where

$$F = \{(x_1, x_2, x_3) \mid x_1 = 1, x_2 = 1\} \text{ and } G = \{(x_1, x_2, x_3) \mid x_1 = 3/2, x_2 = 1/2\}.$$

Let L_1 be the linear ordering of M which occurs by having each element of F precede each element of G, and let L_2 be the linear ordering of M which occurs by having each element of G precede each element of F. Now L_1 and L_2 together realize both $P(M)$ and $P'(M)$. Thus the dimension of each is two. Nevertheless, each vertical plane $x_3 = k$ intersects M in the two elements.

The following two problems are still open:

(1) Characterize those subsets M of E_n for which the λ-dimension of $P(M)$ and the λ-dimension of $P'(M)$ do not exceed n.

(2) Can there exist subsets of $P(E_n)$ and of $P'(E_n)$ which have a finite λ-dimension $> n$?

2. *A*-dimension. A subset E, of power $2^{\aleph_0}$, of the real numbers is said to have property A if no disjoint subsets of E, of power $2^{\aleph_0}$ each, are similar.

The linear order L on S shall be called an A-extension of the partial order P on S if L is an extension of P which is similar to a set having property A.

By the A-dimension of a partial order P on S is meant the minimum number of A-extensions of P which realize P.

There exist partial orders of power $2^{\aleph_0}$ which do not have A-dimension. A trivial example is the order on the set of real numbers. To obtain a condition on P which ensures that it have A-dimension the following lemma is needed.

LEMMA 2.1. *Let G be an ordered sum $G = G_0 + \cdots + G_n$ having property A. Let H be a subset of the real numbers which is an ordered sum $H = H_0 + \cdots + H_n$, where the H_i satisfy the following condition: There exists a permutation f, of the set of integers $\{i \mid i \leqq n\}$ such that $G^*_{f(i)} = H^*_i$, where E^* is the order type of E. Then the set H has property A.*

Proof. Suppose the contrary. Let M and N be two disjoint, similar subsets of H of power $2^{\aleph_0}$ each. Then for some integer s, $H_s \cap M$ is of power $2^{\aleph_0}$. Let Q be a subset of N which is similar to $H_s \cap M$. For some integer t, $H_t \cap Q$ is of power $2^{\aleph_0}$. Let U be a subset of $G_{f(s)}$ which is similar to $H_t \cap Q$ and V a subset of $G_{f(t)}$ which is similar to $H_t \cap Q$. U and V are disjoint since $H_s \cap M$ and $H_t \cap Q$ are disjoint. Thus U and V are two disjoint, similar subsets, of power $2^{\aleph_0}$ each, of G. But this contradicts the fact that G has property A. Consequently no such sets M and N can exist, i. e., H must have property A. q. e. d.

In 4.2 of [6] it was shown that a necessary and sufficient condition for a partial order P on a set E to have λ-dimension was for P to have a λ-extension. Repeating that proof in conjunction with Lemma 2.1 we obtain

THEOREM 2.1. ***A necessary and sufficient condition for a partial order P defined on a set S to have A-dimension is for P to have an A-extension.***

In conclusion we present an example which shows that the λ-dimension need not equal the A-dimension even when both exist. First, however, we state a result which is an immediate consequence of Lemma 3. 1 of [5] and Lemma 3. 1 of [4].

(*) The open interval $(0, 1)$ of the real numbers is not the union of a finite number of sets, each of which either has property A or is of power $< 2^{\aleph_0}$.

A non-empty subset E of the real numbers is said to have property C if, for each element p of E and each positive number δ, each of the two open intervals $(p - \delta, p)$ and $(p, p + \delta)$, of the real number system contains $2^{\aleph_0}$ elements of E.

Let S be a set which has properties A and C. Such sets are known to exist (Theorem 4. 4 of [5]). For some element z_0 of S let $M = \{u \mid u < z_0, u \in S\}$ and $N = \{u \mid u \geqq z_0, u \in S\}$. The set M has properties A and C and thus is similar to a dense subset G of the open interval $E = (0, 1)$ of the real numbers. Let $H = E - G$. By Lemma 3. 1 of [4] H is a dense subset, of power $2^{\aleph_0}$, of E. Define g_0 to be the identity transformation on E. Define g_1 to be a function which is a similarity transformation of G onto M, and which also maps H one to one onto N. The function g_1 is well defined since H and N both are of power $2^{\aleph_0}$. The two functions g_0 and g_1 realize a partial order P on E. Since $g_1(E) = S$ has property A, by Theorem 2. 1, P has A-dimension. The λ-dimension of P is 2. We shall show that the A-dimension of P is infinite.

Suppose that the A-dimension of P is $n + 1$, where n is a positive integer. Let $f_0, f_1, \cdots, f_n$, be $n + 1$ real valued functions of E, such that $f_i(E)$ has property A, which realize P. We now shall proceed to arrive at a contradiction. Note that for any two elements u and v, if $u < v$ in G then $u < v$ in P. For each element x in H let $S_x = \{y \mid y < x, y \in G\}$ and $T_x = \{y \mid y > x, y \in G\}$. Since S_x and T_x are subsets of G, $f_i(S_x) < f_i(T_x)$ for each $i \leqq n$. Two alternatives arise:

(a) There exists an element x in H so that for no integer i is $f_i(S_x) < f_i(x) < f_i(T_x)$. Now in P, $S_x < x$, and x is incomparable with each element of T_x. By hypothesis, for each integer i, an element y_i in T_x can be found so that $f_i(y_i) < f_i(x)$. The set $J = \{y_i \mid i \leqq n\}$, being a subset of T_x, is simply ordered in P. Denote by w the smallest element in J. Clearly $f_i(w) < f_i(x)$ for each i. Thus $w < x$ in P. But this contradicts the fact that x is incomparable with each element in T_x.

(b) For each element x in H, there exists an integer $i(x)$ such that $f_{i(x)}(S_x) < f_{i(x)}(x) < f_{i(x)}(T_x)$. For each $i \leqq n$ let

$$B_i = \{x | x \in H, f_i(S_x) < f_i(x) < f_i(T_x)\}.$$

Notice that B_i is similar to $f_i(B_i)$. Hence B_i either has property A or is of power $< 2^{\aleph_0}$. We conclude that E is the union of a finite family of sets, $\{G\} \cup \{B_i | i \leqq n\}$, each of which has property A or is of power $< 2^{\aleph_0}$. This contradicts (*).

Since (a) and (b) are the only possibilities, and each of these cases cannot occur, we are forced to conclude that the functions f_i cannot exist, i. e., the A-dimension of P is infinite.

THE UNIVERSITY OF MIAMI.

BIBLIOGRAPHY.

[1] Dushnik, B., "Concerning a certain set of arrangements," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 788-796.

[2] Dushnik and Miller, "Partially ordered sets," *American Journal of Mathematics*, vol. 63 (1941), pp. 600-610.

[3] Fuchs, L., "On the extension of the partial order of groups," *American Journal of Mathematics*, vol. 72 (1950), pp. 191-194.

[4] Ginsburg, S., "Some remarks on order types and decompositions of sets," *Transactions of the American Mathematical Society*, vol. 74 (1953), pp. 514-535.

[5] ———, "Further results on order types and decompositions of sets," to appear in the *Transactions of the American Mathematical Society*.

[6] Komm, H., "On the dimension of partially ordered sets," *American Journal of Mathematics*, vol. 70 (1948), pp. 507-520.

[7] Ruziewicz, S., "Sur les fonctions d'une infinité de variables," *Fundamentica Mathematicae*, vol. 26 (1936), pp. 52-55.

[8] Sedmak, V., "Dimension des ensembles partiellement ordonnés associés aux polygones et polyhèdres." *Hrvatsko Prirodoslovno Drŭstvo Glasnik Matematicka-Fizicka i Astronomska Sekcija*, II, 7 (1952), pp. 169-182.

INTERPRETATION OF VECTOR COHOMOLOGY GROUPS.*

By W. H. Cockcroft.

1. Introduction. In a recent note, J. H. C. Whitehead (see [4]; the relevant definitions are repeated in section **2** below) has defined *vector cohomology groups* $\mathfrak{H}_p^n(X, Y; G)$ of groups X and Y over an Abelian group G. The definition generalises the cohomology groups $H^n(Y, G)$ of S. Eilenberg and Saunders Maclane (see [2]), by introducing the group X of operators on Y and G. In his note, J. H. C. Whitehead interpreted the groups $\mathfrak{H}_2^3$ and $\mathfrak{H}_1^2$ in terms of group extensions with operators. His results generalise those of Eilenberg and Maclane which interpret the ordinary cohomology groups $H^3(Y, G)$ and $H^2(Y, G)$ in terms of the group of similarity classes of Y-kernels (with centre G) and the group of (non-operator) group extensions of Y by G respectively (see [2], paras. 4, 5, 6).

The object of this note is to show that the other algebraic interpretations of ordinary cohomology groups can also be generalised. In part I below I will interpret the group $\mathfrak{H}_0^1$ in terms of operator automorphisms of group extensions with operators. The result generalises the well known theorem of Baer (see [1], also [2], para. 4) which relates the group $H^1(Y, G)$ to the group of automorphisms of a group extension of Y by G.

In part II, I define *loops with operators in a group* and show that the Eilenberg and Maclane theory of loop prolongations which is developed in [3] can be generalised to a theory of *loop prolongations with operators* by using the vector cohomology groups. This gives an interpretation of the groups $\mathfrak{H}_{n-1}^n$ for all $n \geq 3$, just as the Eilenberg and Maclane theory gives an interpretation of the groups H^n for all $n \geq 3$.

The result in part I is proved directly. In part II, I use the technique of [4], para. 6 to obtain the results indirectly. This technique is outlined in section **5** below.

Part I.

2. Vector cohomology groups. For convenience I shall reproduce here the definition of the vector cohomology groups of [4]. Let G be an additive Abelian group, which admits a multiplicative group Y as a group of operators.

* Received October 22, 1952.

Let the operators of $y \in Y$ and $g \in G$ give the element $y \cdot g \in G$. Further, let both Y and G admit a multiplicative group X as a group of operators. Let the symbols xg and xy denote the images of $g \in G$ and $y \in Y$ respectively under the operator $x \in X$. Let $xy \cdot g$ denote $(xy) \cdot g$ and let $x(y \cdot g) = (xy) \cdot (xg)$, for all $x \in X$, $y \in Y$ and $g \in G$.

The cochain group $C^n(Y, G)$ of Y over G, $n \geq 0$, admits X as a group of operators from both left and right by the rules:

$$(xc)(y_1, \cdots, y_n) = xc(y_1, \cdots, y_n), \quad (cx)(y_1, \cdots, y_n) = c(xy_1, \cdots, xy_n),$$

where $c \in C^n(Y, G)$. The group Y operates on $C^m(X, G)$, $m \geq 0$, according to the rule

$$(yc)(x_1, \cdots, x_m) = x_1 x_2 \cdots x_m y \cdot c(x_1, \cdots, x_m),$$

where $c \in C^m(X, G)$. It is easily verified that $1c = c$ and that $y(y'c) = (yy')c$ as required.

I recall that the co-boundary operator $\delta: C^n(W, A) \to C^{n+1}(W, A)$, of the group of cochains of a multiplicative group W over an additive Abelian group A, on which W operates from both left and right, is defined by

$$(2.1) \quad (\delta c)(w_1, \cdots, w_{n+1}) = w_1 c(w_2, \cdots, w_{n+1}) + \sum_{i=1}^{n} (-1)^i c(w_1, \cdots, w_i w_{i+1}, \cdots, w_n) + (-1)^{n+1} c(w_1, \cdots, w_n) w_{n+1}.$$

(If W operates only from the left, then we set $cw = c$). Coboundary operators

$$\delta_x: C^m(X, C^n(Y, G)) \to C^{m+1}(X, C^n(Y, G)),$$

$$\delta_y: C^n(Y, C^m(X, G)) \to C^{n+1}(Y, C^m(X, G)),$$

are now defined in exactly this way with respect to the cochain groups $C^m(X, C^n(Y, G))$ and $C^n(Y, C^m(X, G))$. Since any element of either of these groups can be regarded as a function of the form $c(x_1, \cdots, x_m; y_1, \cdots, y_n)$, we identify them and denote them by $C^{m,n}(X, Y; G)$ or simply $C^{m,n}$. It is trivial that $\delta_x \delta_x = 0$, $\delta_y \delta_y = 0$ and that $\delta_x \delta_y = \delta_y \delta_x$.

Let $C^n = C^n(X, Y; G)$ be the group, with vector addition, in which the elements are all ordered sets $(c_0, c_1, \cdots, c_n)$, where $c_k \in C^{k,n-k}(X, Y; G)$. The elements $(c_0, \cdots, c_n)$ of this group are *vector cochains*. A homomorphism $\Delta: C^n \to C^{n+1}$ is defined by

$$\Delta(c_0, \cdots, c_n) = (c'_0, \cdots, c'_{n+1}), \qquad c'_k = \delta_x c_{k-1} + (-1)^k \delta_y c_k,$$

with $c_{-1} = c_{n+1} = 0$. It follows from the fact that $\delta_x \delta_x = 0$, $\delta_y \delta_y = 0$ and that $\delta_x \delta_y = \delta_y \delta_x$ that $\Delta\Delta = 0$.

Let $0 \leq p \leq n$, and let $\mathfrak{C}_p^n \subset C^n$ denote the subgroup of C^n which consists of the vector cochains $(c_0, \cdots, c_n)$ such that $c_{p+1} = 0, \cdots, c_n = 0$. These are denoted by $(c_0, \cdots, c_p)^n$. When $p = -1$, the group $\mathfrak{C}_{-1}^0$ consists of the cochains $(g)^0$, for all elements $g \in G$ such that $xg = g$ for all $x \in X$. Note that since $xg = g$, we have $\delta_x g = 0$ and therefore $\Delta(g)^0 = (\delta_y g)^1$. The vector cohomology group $\mathfrak{H}_p^n$ is then defined by $\mathfrak{H}_p^n = \mathfrak{Z}_p^n - \Delta\mathfrak{C}^{n-1}$, where $\mathfrak{Z}_p^n = \mathfrak{C}_p^n \cap \Delta^{-1}(0)$ is the group of *vector cocycles.*

3. Statement of Theorem. Let (E, ϕ) be an X-extension of an Abelian group G by a group Y, i. e. E and Y admit X as a group of operators and there is an operator homomorphism $\phi: E \to Y$, onto Y, with kernel G. Since E admits X as a group of operators, and ϕ is an operator homomorphism, G also admits X as a group of operators, with $x: G \approx G$ the automorphism induced by $x: E \approx E$, for all $x \in X$.

Let (E', ϕ') be similarly defined in terms of groups X', Y' and G'. Both E and E' will be written additively though in general neither will be Abelian.

Suppose $f: X \to X'$ is a given homomorphism. A homomorphism $\alpha: E \to E'$ is an *operator homomorphism associated with* f if and only if $\alpha(xe) = (fx)\alpha e$ for all $x \in X$, $e \in E$. An operator homomorphism $\alpha: E \to E'$ is a *proper homomorphism* if and only if $\alpha G \subset G'$. Thus a proper homomorphism $\alpha: E \to E'$ *induces* an operator homomorphism $\theta = \alpha | G: G \to G'$ associated with f in the sense that $\theta(xg) = (fx)\theta g$, and an operator homomorphism $\psi: Y \to Y'$ defined by $\psi\phi = \phi'\alpha$, also associated with f, i. e. such that $\psi(xy) = (fx)\psi y$.

Let representatives $e(y) \in \phi^{-1}(y)$, $e'(y') \in \phi'^{-1}(y')$ be chosen for all $y \in Y$, $y' \in Y'$, with $e(1) = 0$, $e'(1) = 0$. Let Y operate on G and Y' on G' by the usual rules:

$$y \cdot g = e(y) + g - e(y) \qquad (y \in Y; g \in G),$$

and

$$y' \cdot g' = e'(y') + g' - e'(y') \qquad (y' \in Y'; g' \in G').$$

Finally the operation of Y on G' is defined by $y \cdot g' = (\psi y) \cdot g'$ for all $y \in Y$, $g' \in G'$.

It is now a trivial matter to verify that, with these definitions of operations, the groups X, Y, G' satisfy all the requirements set out in section 2 above. Thus we may define $\mathfrak{H}_p^n(X, Y; G')$. I prove the

THEOREM. *The proper homomorphisms $\alpha: E \to E'$, associated with $f: X \to X'$, all of which induce the same homomorphisms $\theta: G \to G'$, $\psi: Y \to Y'$ are in a one-one correspondence with the elements of $\mathfrak{Z}_0^1(X, Y; G')$. In this*

correspondence the elements of $\Delta\mathfrak{C}_{-1}{}^0(X, Y; G')$ *correspond to proper homomorphisms* $\bar{\alpha}: E \to E'$ *of the type*

$$\bar{\alpha}e = g' + \alpha e - g' \qquad (e \in E),$$

where $(g')^0 \in \mathfrak{C}_{-1}{}^0$, *and where* $\alpha: E \to E'$ *is, as above, a proper homomorphism associated with* f, *and inducing* θ *and* ψ.

If $(E', \phi') = (E, \phi)$, $X' = X$, $Y' = Y$ and $f = 1: X \to X$, $\theta = 1: G \to G$, $\psi = 1: Y \to Y$, where $1: Q \to Q$ denotes the identity automorphism of any group Q onto itself, we may define the group A_1 of proper automorphisms of E, associated with $1: X \to X$, which induce $1: G \to G$ and $1: Y \to Y$. In this case the one-one correspondence of the theorem becomes an isomorphism, and if A_2 denotes the subgroup of A_1 which consists of automorphisms $e \to g + e - g$, where $xg = g$, we have the

COROLLARY. $A_1/A_2 \approx \mathfrak{H}_0{}^1(X, Y; G')$.

If $X = 1$, this reduces immediately to the theorem of Baer.

4. Proof of Theorem. I first recall that if we are given the representatives $e(y)$, $e'(y')$, chosen in section **3** above, we have associated with the extensions (E, ϕ) and (E', ϕ') the cochains:

$$c_1(x; y) = xe(y) - e(xy) \qquad (x \in X; y \in Y),$$
$$c_0(y_1, y_2) = e(y_1) + e(y_2) - e(y_1y_2) \qquad (y_1, y_2 \in Y),$$
$$c'_1(x'; y') = x'e'(y') - e'(x'y') \qquad (x' \in X', y' \in Y'),$$

and

$$c'_0(y'_1, y'_2) = e'(y'_1) + e'(y'_2) - e'(y'_1y'_2) \qquad (y'_1, y'_2 \in Y').$$

It is a simple matter to verify that (see [4], p. 224) $\delta_y c_0 = 0$, $\delta_x c_1 = 0$, $\delta_y c'_0 = 0$, $\delta_x c'_1 = 0$, $\delta_x c_0 = \delta_y c_1$, and $\delta_x c'_0 = \delta_y c'_1$.

Now let $\alpha_1: E \to E'$ be any proper homomorphism which is associated with a homomorphism $f: X \to X'$ and which induces $\theta: G \to G'$ and $\psi: Y \to Y'$. Any element of E may be put in the form $g + e(y)$ for some $g \in G$ and $y \in Y$, and similarly for E'. Suppose $\alpha_1(g + e(y)) = g' + e'(y')$ for some $g' \in G'$, $y' \in Y'$. Since α_1 induces θ and ψ we have $g' + e'(y') = \theta g + a_1(y) + e'(\psi y)$ where $a_1(y) \in C^{0,1}(X, Y; G')$. Now if $g_1, g_2 \in G$ and $y_1, y_2 \in Y$, then

$$\begin{aligned}
\alpha_1(g_1 + e(y_1) + g_2 + e(y_2)) &= \alpha_1(g_1 + y_1 \cdot g_2 + c_0(y_1, y_2) + e(y_1y_2)) \\
&= \theta g_1 + \theta(y_1 \cdot g_2) + \theta c_0(y_1, y_2) + a_1(y_1y_2) + e'(\psi y_1y_2) \\
&= a_1(y_1y_2) - a_1(y_1) - (\psi y_1) \cdot a_1(y_2) + \theta c_0(y_1, y_2) \\
&\quad - c'_0(\psi y_1, \psi y_2) + \theta g_1 + a_1(y_1) + (\psi y_1) \cdot \theta g_2 \\
&\quad + (\psi y_1) \cdot a_1(y_2) + c'_0(\psi y_1, \psi y_2) + e'(\psi y_1 y_2).
\end{aligned}$$

Therefore

$$(4.1)\quad \begin{aligned}\alpha_1(g_1 + e(y_1) + g_2 + e(y_2)) &= -(\delta_y a_1)(y_1, y_2) + \theta c_0(y_1, y_2) - c'_0(\psi y_1, \psi y_2)\\ &\quad + \theta g_1 + a_1(y_1) + e'(\psi y_1) + \theta g_2 + a_1(y_2) + e'(\psi y_2).\end{aligned}$$

But by definition, α_1 is a homomorphism; thus

$$(4.2)\quad \begin{aligned}\alpha_1(g_1 + e(y_1) + g_2 + e(y_2)) &= \alpha_1(g_1 + e(y_1)) + \alpha_1(g_2 + e(y_2))\\ &= \theta g_1 + a_1(y_1) + e'(\psi y_1) + \theta g_2 + a_1(y_2) + e'(\psi y_2),\end{aligned}$$

and therefore, comparing this with (4.1), we have

$$(4.3)\qquad (\delta_y a_1)(y_1, y_2) = \theta c_0(y_1, y_2) - c'_0(\psi y_1, \psi y_2).$$

Again, if $g \in G$, $x \in X$, $y \in Y$, we have

$$\begin{aligned}\alpha_1(xg + xe(y)) &= \alpha_1(xg + c_1(x; y) + e(xy))\\ &= \theta xg + \theta c_1(x; y) + a_1(xy) + e'(\psi(xy))\\ &= a_1(xy) - (fx)a_1(y) + \theta c_1(x; y) - c'_1(fx; \psi y)\\ &\quad + (fx)(\theta g + a_1(y) + e'(\psi y)).\end{aligned}$$

Therefore

$$(4.4)\quad \begin{aligned}\alpha_1(xg + xe(y)) &= -(\delta_x a_1)(x; y) + \theta c_1(x; y) - c'_1(fx; \psi y)\\ &\quad + (fx)(\alpha_1(g + e(y))).\end{aligned}$$

But by definition, α_1 is an operator homomorphism; thus

$$(4.5)\quad \alpha_1(xg + xe(y)) = \alpha_1\{x(g + e(y))\} = (fx)(\alpha_1(g + e(y))),$$

and therefore, comparing this with (4.3), we have

$$(4.6)\qquad (\delta_x a_1)(x; y) = \theta c_1(x; y) - c'_1(fx; \psi y).$$

Conversely, if $\theta: G \to G'$ and $\psi: Y \to Y'$ are operator homomorphisms associated with $f: X \to X'$, we may define a mapping $\alpha_1: E \to E'$ by setting

$$\alpha_1(g + e(y)) = \theta g + a_1(y) + e'(\psi y),$$

for all $g \in G$ and $y \in Y$, where $a_1 \in C^{0,1}(X, Y; G')$ and satisfies conditions (4.3) and (4.6) above. It is a straightforward matter to prove that α_1, so defined, is a proper homomorphism associated with $f: X \to X'$ and inducing θ and ψ. Firstly it is obvious that α_1 induces θ and ψ as required. Secondly, it is a homomorphism, since (4.1) and (4.3) imply (4.2). Thirdly, it is an operator homomorphism associated with $f: X \to X'$, since (4.4) and (4.6)

imply (4.5). Finally it is obvious that $\alpha G \subset G'$ and α_1 is therefore a proper homomorphism of the required type.

Let $\alpha_2 : E \to E'$ be another proper homomorphism associated with f and inducing θ and ψ. Further let $\alpha_2(g + e(y)) = \theta g + a_2(y) + e'(\psi y)$ $(g \in G; y \in Y)$. Again we have $a_2(y) \in C^{0,1}(X, Y; G')$, and $(\delta_y a_2)(y_1, y_2)$ and $(\delta_x a_1)(x; y)$ satisfy similar conditions to (4.3) and (4.6). Thus $(a_1 - a_2)^1$ is a vector cochain in the group $\mathfrak{C}_0^1(X, Y; G')$ and it follows from (4.3) and (4.6) that $\Delta(a_1 - a_2)^1 = 0$. That is, $(a_1 - a_2)^1$ is a vector cocycle in $\mathfrak{Z}_0^1(X, Y; G')$. It is obviously independent of the choice of the representatives $e(y)$ and $e'(y')$.

Suppose now that $\alpha_1 : E \to E'$ is fixed and let $\alpha_2 : E \to E'$ be *any* proper homomorphism associated with f and inducing θ and ψ. I say that the correspondence $\alpha_2 \to (a_1 - a_2)^1$ is one-one with $\mathfrak{Z}_0^1(X, Y; G')$. For suppose $\alpha_2, \alpha'_2 : E \to E'$ are proper homomorphisms associated with f and inducing θ and ψ. If α_2 and α'_2 have associated with them, as above, the functions $a_2(y)$ and $a'_2(y)$ respectively, and if $(a_1 - a_2)^1 = (a_1 - a'_2)^1$, then $a_1 - a_2 = a_1 - a'_2$. Hence $a_2 = a'_2$ and $\alpha_2 = \alpha'_2 : E \to E'$. Thus the correspondence is one-one. It is onto $\mathfrak{Z}_0^1(X, Y; G')$ for if $(d_0)^1$ is a cocycle in $\mathfrak{Z}_0^1(X, Y; G')$, define $\alpha_2 : E \to E'$ by setting

$$\alpha_2(g + e(y)) = \theta g + a_2(y) + e'(\psi y),$$

for all $g \in G$, $y \in Y$, where $a_2(y) = a_1(y) - d_0(y)$. Then

$$(\delta_y a_2)(y_1, y_2) = \theta c_0(y_1, y_2) - c'_0(\psi y_1, \psi y_2),$$

since $\Delta(d_0)^1 = 0$ implies that $(\delta_y d_0)(y_1, y_2) = 0$. Similarly

$$(\delta_x a_2)(x; y) = \theta c_1(x; y) - c'_1(fx; \psi y),$$

and it follows that $\alpha_2 : E \to E'$ is a proper homomorphism of the required type, which by definition corresponds to the cocycle $(d_0)^1$. This establishes the first part of the theorem. Finally, suppose $(g')^0 \in \mathfrak{C}_{-1}^0$. Define $\alpha_2 : E \to E'$ by setting $\alpha_2(g + e(y)) = g' + \alpha_1(g + e(y)) - g'$, for all $g \in G$ and $y \in Y$. Thus

$$\begin{aligned} \alpha_2(g + e(y)) &= \theta g + a_1(y) + g' + e'(\psi y) - g' \\ &= \theta g + a_1(y) + g' - y \cdot g' + e'(\psi y) \\ &= \theta g + a_2(y) + e'(\psi y) \end{aligned}$$

where $a_2(y) = a_1(y) - (\delta_y g')(y)$. Obviously

$$(\delta_y a_2)(y_1, y_2) = (\delta_y a_1)(y_1, y_2) \qquad (y_1, y_2 \in Y),$$

and

$$(\delta_x a_2)(x; y) = (\delta_x a_1)(x; y) \qquad (x \in X; y \in Y).$$

It follows immediately that α_2, so defined, is a proper homomorphism associated with $f: X \to X'$ and inducing θ and ψ. Moreover α_2 corresponds to the cocycle $(a_1 - a_2)^1 = (\delta_y g')^1$, and $(\delta_y g')^1 \in \Delta\mathfrak{C}_{-1}{}^0(X, Y; G')$. Thus the Theorem is proved.

Suppose now that the conditions of the Corollary hold. It is obvious that the proper homomorphisms associated with $1: X \to X$, which induce the identity automorphisms $1: G \to G$, $1: Y \to Y$, form a multiplicative group A_1 in which the product of two homomorphisms α, β is the homomorphism $\beta\alpha: E \to E$. Moreover A_1 is Abelian. Set $\alpha_1 = 1: E \to E$ in the proof of the Theorem and suppose $\alpha_i, \alpha_j: E \to E$ to be proper automorphisms associated with $1: X \to X$, given by $\alpha_r(g + e(y)) = g + a_r(y) + e(y)$, $(r = i, j)$. Thus α_i and α_j correspond to the cocycles $(-a_i)^1$ and $(-a_j)^1$ respectively. But $\alpha_j\alpha_i(g + e(y)) = g + a_i(y) + a_j(y) + e(y)$. Thus $\alpha_j\alpha_i: E \to E$ corresponds to the cocycle $(-a_j - a_i)^1 \in \mathfrak{Z}_0{}^1(X, Y; G)$, and the one-one correspondence of the theorem is in this case an isomorphism onto $\mathfrak{Z}_0{}^1(X, Y; G)$. Under this isomorphism $\Delta\mathfrak{C}_{-1}{}^0(X, Y; G)$ obviously corresponds to the subgroup A_2 of A_1 which consists of all automorphisms, $e \to g + e - g$ where $(g)^0 \in \mathfrak{C}_{-1}{}^0(X, Y; G)$. Thus the Corollary follows.

Part II.

5. Outline of method. For convenience I shall first recall the definitions of loop prolongations, and the classification theorems proved by Eilenberg and Maclane (cf. [2], para. 17, and [3]).

A *loop* L is essentially a non-associative multiplicative group; its elements satisfy the following axioms:

(5.1) for $a, b \in L$, the element $ab \in L$ is uniquely defined;

(5.2) there is a unit element $1 \in L$ such that $a1 = a = 1a$ for all $a \in L$;

(5.3) for any $a, b \in L$ the equations $ax = b$ and $ya = b$ have unique solutions $x, y \in L$.

Given $a, b \in L$, the associator $A(a, b, c) \in L$ is defined by the equation $a(bc) = A(a, b, c)[(ab)c]$. Higher associators are defined inductively by

$$A(a_1, \cdots, a_{2n}, a) = A(a_1, a_2, A(a_3, \cdots, a_{2n}, a)).$$

A *loop prolongation* of a group Y by an Abelian group G, with given operators of Y on G, is a pair (L, ϕ) satisfying

(5.4) L is a loop containing G as a subgroup;

(5.5) ϕ is a homomorphism of L onto Y with kernel a group K containing G;

(5.6) $A(k, a, b) = 1 = A(a, k, b)$ for $a, b \in L$, $k \in K$;

(5.7) $kA = Ak$ for any $k \in K$ and any associator $A = A(a, b, c)$;

(5.8) $ag = (\phi(a)g)a$ for $a \in L$, $g \in G$.

It is usual in the theory of loop prolongations to use the notation $a \cdot b$ for the element z which is the unique solution of the equation $ab = za$. I shall do this and will drop the notation $y \cdot g$ for the operation of an element y of a group Y on $g \in G$, which was used in part I above. This will be written simply yg in future. No ambiguity will arise in what follows because of this. Thus formula (5.8) now reads $a \cdot g = \phi(a)g$.

In [3], Eilenberg and Maclane define two prolongations (L_1, ϕ_1) and (L_2, ϕ_2) to be *equivalent* if and only if there is an isomorphism $\theta: L_1 \approx L_2$ such that $\phi_2\theta = \phi_1: L_1 \to Y$. They also define a multiplication $(L_1, \phi_1) \otimes (L_2, \phi_2)$ of prolongations by considering the loop P of pairs (a_1, a_2), $a_1 \in L_1$, $a_2 \in L_2$ such that $\phi_1 a_1 = \phi_2 a_2$. The loop $L_1 \otimes L_2$ is defined to be the factor loop P/U where U is the set of all pairs (g, g^{-1}) for $g \in G$. These definitions follow those used in group extension theory (see [1], para. 4). The prolongations are now divided into four classes defined by the conditions

$$O_n^L: A(a_1, \cdots, a_{2n}, a) = 1, \qquad O_n^K: A(a_1, \cdots, a_{2n}, k) = 1,$$
$$G_n^L: A(a_1, \cdots, a_{2n}, a) \in G, \qquad G_n^K: A(a_1, \cdots, a_{2n}, k) \in G,$$

where $n > 0$, $a_1, \cdots, a_{2n}$, $a \in L$, and $k \in K$. A further class R_n is defined which consists of those prolongations (L, ϕ) of Y by G which contain a prolongation (L', ϕ') of Y by G which is in class O_n^K (i.e. $L' \subset L$, $\phi | L' = \phi'$ and $(L', \phi') \in O_n^K$). Finally we have a class G_0^K defined by the condition $K = G$.

With each loop prolongation is associated a characteristic cochain $g_u{}^r \in C^r(Y, K)$ with values in the centre Z of K for $r > 2$, defined by taking representatives $u(y) \in \phi^{-1}(y)$ for all $y \in Y$, with $u(1) = 1$, and by setting

$$u(y_1)u(y_2) = g_u{}^2(y_1, y_2)u(y_1y_2),$$
$$g_u{}^{2n+1}(y_1, \cdots, y_{2n}, y) = A(u(y_1), \cdots, u(y_{2n}), u(y)) \qquad (n \geq 1),$$
$$g_u{}^{2n+2}(y_1, \cdots, y_{2n}, y, y') = A(u(y_1), \cdots, u(y_{2n}), g_u{}^2(y, y')) \qquad (n \geq 1),$$

for all $y_1, \cdots, y_{2n}, y, y' \in Y$. These cochains give rise to homomorphisms:

$$g^{2n+1}: O_n^K \cap G_n^L \to Z^{2n+1}(Y, G), \qquad g^{2n+2}: O_{n+1}{}^L \cap G_n^K \to H^{2n+2}(Y, G)$$

for $n > 0$. After defining the products of classes, e. g. ${O_n}^L \otimes {G_{n-1}}^K$, as the set of prolongations which are products of prolongations in the classes, Eilenberg and Maclane prove that these homomorphisms induce the isomorphisms in the following theorems.

THEOREM EM1. $${O_n}^K \cap {G_n}^L/{O_n}^L \otimes {G_{n-1}}^K \approx H^{2n+1}(Y, G) \qquad (n > 0),$$

THEOREM EM2. $${O_{n+1}}^L \cap {G_n}^K/R_n \otimes {G_n}^L \approx H^{2n+2}(Y, G) \qquad (n > 0).$$

In each case the system on the left is a group of modulo classes.

These then are the classification theorems of Eilenberg and Maclane which are proved in [3]. In section **6** below I define *loops with operators* and in section **7** *X-prolongations* (L, ϕ) of a group Y by a group G are defined. These are essentially loops L with operators in X such that $\phi: L \to Y$ is an operator homomorphism whose kernel K is a group which contains the group G.

The classification of such prolongations could be done directly, defining characteristic vector cochains associated with X-prolongations and following the pattern of [3]. However it is simpler to use the indirect technique which J. H. C. Whitehead employs in [4]. To this end I first define the group W which is a group extension of X by Y. Its elements are all pairs (y, x) with $(y_1, x_1)(y_2, x_2) = (y_1 x_1 y_2, x_1 x_2)$ $(x_i \in X; y_i \in Y)$. For convenience y is identified with $(y, 1)$ and $\bar{x}$ denotes $(1, x)$. Thus $(y, x) = y\bar{x}$, and w, w_i are used to denote $y\bar{x}$ and $y_i\bar{x}_i$. The operation of W on G is defined by setting $wg = y(xg)$, for all $w \in W$, $g \in G$.

I now associate with each X-prolongation (L, ϕ) an ordinary (i. e. non-operator) prolongation (L^*, ϕ^*) of the group W by the group G. The correspondence $(L, \phi) \to (L^*, \phi^*)$ is shown to be one-one to within an equivalence. Furthermore, the classes of X-prolongations which are defined, ${O_n}^{*L}$, ${O_n}^{*K}$, ${G_n}^{*L}$, ${G_n}^{*K}$ are carried over into the corresponding classes ${O_n}^L$, ${O_n}^K$, etc. by the correspondence. Products of X-prolongations and their classes are defined and are preserved by the correspondence in a similar way. Thus the classification of X-prolongations is referred to the classification of prolongations of W by G, which are called *associated prolongations.*

In order to use the theory of [3] to classify the associated prolongations we require a new cohomology group $H_*^n(W, G)$ which is defined as follows. Given W, as above, let $\lambda: X \to W$ be the homomorphism defined by setting $\lambda x = \bar{x}$, and let $\lambda^*: C^n(W, G) \to C^n(X, G)$ be the homomorphism induced by λ. Denote $\lambda^{*-1}(0)$ by ${C_*}^n$, and let δ be the coboundary operator defined by (2. 1). Since $\bar{x}g = xg$, it follows that $\lambda^*\delta = \delta\lambda^*$, and hence $\delta {C_*}^n \subset {C_*}^{n+1}$.

Let $\delta_* : C_*{}^n \to C_*{}^{n+1}$ be the homomorphism induced by δ and let $H_*{}^0, H_*{}^1, \cdots$, be the cohomology groups which are defined in terms of $C_*{}^0, C_*{}^1, \cdots$, and the coboundary operator δ_*. That these groups are the ones required for our classification, follows from the fact that the characteristic cochains of the associated prolongations belong to $C_*{}^n(W, K)$ and not simply to $C^n(W, K)$, as do the characteristic cochains of [3]. The results of [3] can obviously be taken over as they stand, setting $C_*{}^n$, $B_*{}^n$, $Z_*{}^n$, $H_*{}^n$ for the corresponding C^n, B^n, Z^n and H^n (I use the normal notation for cochain, cocycle, coboundary and cohomology groups) except for one minor and obvious alteration: it is necessary to prove a new existence theorem, replacing Theorem 12.1 of [3], which proves that given a cocycle εZ^n, there is a prolongation with that cocycle as its characteristic cocycle. Here we require that the constructed prolongation should not be just any prolongation but an *associated* one. The proof I give is a modification of Theorem 12.1 of [3], and after certain initial differences, which are given in full, it is identical with that in [3]. I therefore omit the identical portions and prove that, having constructed a prolongation (L^*, ϕ^*) as in [3], (L^*, ϕ^*) is an associated prolongation. We are then in a position to quote the classification theorems of [3], EM1 and EM2 above, to classify the associated prolongations in terms of the groups $H_*{}^n$.

Now it is proved in [4], that $H_*{}^n \approx \mathfrak{H}_{n-1}{}^n$ for each $n = 0, 1, 2, \cdots$. Thus the classification of associated prolongations by means of the groups $H_*{}^n$, together with this isomorphism and the one-one correspondence of the associated prolongations with the X-prolongations yields a classification of X-prolongations in terms of the groups $\mathfrak{H}_{n-1}{}^n$.

In a final section I indicate the definition of a characteristic vector cochain of an X-prolongation which would yield these results directly.

6. Loops with operators. Let X be a multiplicative group and L a loop. The loop L will be said to have *operators in the group* X, or to be operated on by X, if given $x \varepsilon X$ and $a \varepsilon L$ there is a unique element $xa \varepsilon L$ such that $x1 = 1$, $1a = a$, where the symbol 1 is of course used for both the unit of L and of X.

The *vector*, or *operator*, *associators* of L are triples $(A_0{}^3, A_1{}^3, A_2{}^3)$ of elements of L which are uniquely defined by the equations:

$$a_1(a_2a_3) = A_0{}^3(a_1, a_2, a_3)\{(a_1a_2)a_3\} \qquad (a_1, a_2, a_3 \varepsilon L);$$

$$x_1(a_1a_2) = A_1{}^3(x_1; a_1, a_2)\{(x_1a_1)(x_2a_2)\} \qquad (x_1 \varepsilon X; a_1, a_2 \varepsilon L);$$

$$x_1(x_2a_1) = A_2{}^3(x_1, x_2; a_1)\{(x_1x_2)a\} \qquad (x_1, x_2 \varepsilon X; a_1 \varepsilon L).$$

Higher order vector associators are defined inductively. Thus the vector associator $(A_0{}^{2n+1}, \cdots, A_{2n}{}^{2n+1})$ is defined by

$$A_0{}^{2n+1}(a_1, \cdots, a_{2n+1}) = A_0{}^3(a_1, a_2, A_0{}^{2n-1}(a_3, \cdots, a_{2n+1}))\,;$$

$$A_1{}^{2n+1}(x_1\,; a_1, \cdots, a_{2n}) = A_1{}^3(x_1\,; a_1, A_0{}^{2n-2}(a_2, \cdots, a_{2n}))\,;$$

$$A_r{}^{2n+1}(x_1, \cdots, x_r\,; a_1, \cdots, a_{2n+1-r}) = A_2{}^3(x_1, x_2\,; A_{r-2}{}^{2n-1}(x_3, \cdots, x_r\,; a_1, \cdots, a_{2n+1-r})$$

where $2 \leq r \leq 2n$ and $a_1, \cdots, a_{2n+1} \in L, x_1, \cdots, x_r \in X$.

A subgroup S of L will be said to be an *operator* subgroup if $xS \subset S$ for all $x \in X$. If also $A_0{}^3(s, a, b) = 1 = A_0{}^3(a, s, b)$, $A_1{}^3(x\,; s, a) = 1$, for all $s \in S$, $a, b \in L$ and $x \in X$, then we can define in the usual way (see e. g. [3], para. 3, and the references cited there) the factor loop L/S of cosets Sa, of S in L. Furthermore, this loop will have operators in X defined by $x(Sa) = S(xa)$ and the natural homomorphism $\mu : L \to L/S$ defined by $\mu a = Sa$, will be an operator homomorphism in the sense that $\mu x = x\mu$. The loop L/S is obviously associative (and hence a group) if, and only if, $A_0{}^3(a_1, a_2, a_3) \in S$ for all $a_1, a_2, a_3 \in L$. If also $A_1{}^3(x_1\,; a_1, a_2)$ and $A_2{}^3(x_1, x_2\,; a_1) \in S$ for all $x_1, x_2 \in X$ and $a_1, a_2 \in L$, then L/S is obviously a group which admits X as a group of operators in the usual sense.

7. Prolongations with operators. As in part I, section **2**, above, let X be a multiplicative group acting as a group of operators on a multiplicative group Y, and furthermore let both X and Y operate on an Abelian group G subject to the usual condition, now written $x(yg) = (xy)(xg)$, for all $x \in X$, $y \in Y$, $g \in G$. An X-prolongation of Y by G is a pair (L, ϕ) satisfying the conditions:

(7. 1) L is a loop with operators in X and containing G as an operator subgroup;

(7. 2) ϕ is an operator homomorphism of L onto Y with kernel a subgroup K of L which contains G;

(7. 3) $A_0{}^3(k, a, b) = 1 = A_0{}^3(a, k, b)$ $\qquad (k \in K\,; a, b \in L)$;

(7. 4) $A_1{}^3(x\,; k, a) = 1$ $\qquad (x \in X\,; k \in K\,; a \in L)$;

(7. 5) $kA_i{}^3 = A_i{}^3 k$ $\qquad (k \in K\,; i = 0, 1, 2)$;

(7. 6) $a \cdot g = (\phi a)g$ $\qquad (a \in L\,; g \in G)$;

(7. 7) $x(a \cdot g) = (xa) \cdot (xg)$ $\qquad (x \in X\,; a \in L\,; g \in G)$.

Here, as stated before, the symbol $a \cdot g$ denotes the solution z of the equation $ag = za$.

Two such X-prolongations of Y by G, (L_1, ϕ_1), (L_2, ϕ_2), will be said to be *equivalent* if there is an isomorphism $\theta: L_1 \to L_2$ such that $x\theta = \theta x$, $\phi_1 = \phi_2\theta$. A multiplication $(L_1, \phi_1) \otimes (L_2, \phi_2)$ is defined by considering the loop P of pairs (a_1, a_2), $a_1 \in L_1$, $a_2 \in L_2$ such that $\phi_1 a_1 = \phi_2 a_2$. Multiplication is defined by setting $(a_1, a_2)(b_1, b_2) = (a_1 a_2, b_1 b_2)$. The product loop $L_1 \otimes L_2$ is then defined as P/U, where U is the set of pairs (g, g^{-1}), $g \in G$. The operation of X on $L_1 \otimes L_2$ is then defined in the obvious way, i. e. $x(a_1, a_2) = (xa_1, xa_2)$. A homomorphism $\phi: L_1 \otimes L_2 \to Y$ is defined by setting $\phi[a_1, a_2] = \phi_1(a_1) = \phi_2(a_2)$, for all elements $[a_1, a_2]$ of $L_1 \otimes L_2$, and the product is thus an X-prolongation of Y by G.

8. The associated prolongation. Let W be the group extension of X by Y which is defined in section **6** above. Let (L, ϕ) be an X-prolongation of Y by G. Consider the set L^* of elements (a, x) for all $a \in L$, $x \in X$, with a multiplication defined by

$$(a_1, x_1)(a_2, x_2) = (a_1(x_1 a_2), x_1 x_2) \qquad (a_1, a_2 \in L; x_1, x_2 \in X).$$

LEMMA 8a. *L^* is a loop.*

The proof which will be omitted, is trivial and consists merely of showing that L^* satisfies the conditions (5. 1), (5. 2), (5. 3) above.

A mapping $\phi^*: L^* \to W$ is defined by setting $\phi^*(a, x) = (\phi a)\bar{x}$ for all $(a, x) \in L^*$.

LEMMA 8b. *ϕ^* is a homomorphism.*

The proof is trivial and will be omitted.

Let us for convenience identify the elements $a \in L$ with the elements $(a, 1) \in L^*$, and denote $(1, x)$ by x.

LEMMA 8c. *(L^*, ϕ^*) is a prolongation of W by G.*

We require to prove that (L^*, ϕ^*) satisfies the conditions (5. 4)-(5. 8) above.

With the identification proposed above, L^* obviously contains L and therefore it contains G as a subgroup. This is (5. 4). Also $\phi^*: L^* \to W$ is obviously onto W and has kernel $\phi^{*-1}(1) = \phi^{-1}(1) = K$, which contains G. This is (5. 5).

Let the (non-operator) associators of L^* be written $A^*(a_1 x_1, a_2 x_2, a_3 x_3)$ for all $a_1, a_2, a_3 \in L$, $x_1, x_2, x_3 \in X$. From the fact that

$$A^*(a_1 x_1, a_2 x_2, a_3 x_3)\{[(a_1 x_1)(a_2 x_2)](a_3 x_3)\} = (a_1 x_1)[(a_2 x_2)(a_3 x_3)]$$

it follows that

$$A^*(a_1\boldsymbol{x}_1, a_2\boldsymbol{x}_2, a_3\boldsymbol{x}_3)\{([a_1(x_1a_2)][(x_1x_2)a_3])\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3\} = \{a_1[x_1(a_2(x_2a_3))]\}\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3.$$

Manipulation of the product $\{a_1[x_1(a_2(x_2a_3))]\}$ brings it to the form $(\alpha)\{[a_1(x_1a_2)][(x_1x_2)a_3]\}$ where, using the additive notation for the centre Z of K,

$$\alpha = A_0{}^3(a_1, x_1a_2, x_1(x_2a_3)) + a_1 \cdot A_1{}^3(x_1; a_2, x_2a_3) + [a_1(x_1a_2)] \cdot A_2{}^3(x_1, x_2; a_3).$$

Thus

$$A^*(a_1\boldsymbol{x}_1, a_2\boldsymbol{x}_2, a_3\boldsymbol{x}_3) = A_0{}^3(a_1, x_1a_2, x_1(x_2a_3)) + a_1 \cdot A_1{}^3(x_1; a_2, x_2a_3) + [a_1(x_1a_2)] \cdot A_2{}^3(x_1, x_2; a_3).$$

Using (7.3), (7.4) and (7.5) above, it follows that

$$A^*(k, a_2\boldsymbol{x}_2, a_3\boldsymbol{x}_3) = 1 = A^*(a_1\boldsymbol{x}_1, k, a_3\boldsymbol{x}_3) \qquad (k \in K),$$

and $A^*k = kA^*$, for any associator A^*. This proves (5.6) and (5.7).

Finally, if $a \in L$, $x \in X$ **and** $g \in G$,

$$\begin{aligned}\{(\phi^*(a\boldsymbol{x}))g\}(a\boldsymbol{x}) &= \{(\phi a)(xg)\}(a\boldsymbol{x}) \\ &= (a \cdot xg)(a\boldsymbol{x}) && \text{(by (7.6))} \\ &= \{(a \cdot xg)a\}\boldsymbol{x} && \text{(by (5.6))} \\ &= \{a(xg)\}\boldsymbol{x} \\ &= \{(a\boldsymbol{x}) \cdot g\}(a\boldsymbol{x}) && \text{(by (5.6)).}\end{aligned}$$

Thus $(\phi^*(a\boldsymbol{x}))g = (a\boldsymbol{x}) \cdot g$ as required by (5.8), and the proof of the lemma is complete.

This prolongation (L^*, ϕ^*) is the *prolongation associated with* (L, ϕ). Notice that, by definition of the multiplication of L^*, if $u^*(w) = u(y)\boldsymbol{x}$, where $\phi u(y) = y$, $u(1) = 1$, then $\phi^*u^*(w) = w$, $u^*(1) = 1$, and

$$u^*(\bar{x}_1)u^*(\bar{x}_2) = u^*(\bar{x}_1\bar{x}_2) \qquad (x_1, x_2 \in X),$$

$$A^*(a_1\boldsymbol{x}_1, a_2, u^*(\bar{x}_2)) = 1 = A^*(a_1, u^*(\bar{x}_1), a_2\boldsymbol{x}_2) \quad (x_1, x_2 \in X; a_1, a_2 \in L),$$

$$A^*(a_1\boldsymbol{x}_1, u^*(\bar{x}_2), u^*(\bar{x}_3)) = 1 \qquad (x_1, x_2, x_3 \in X; a_1 \in L).$$

It is these properties which characterise an *associated prolongation* (L^*, ϕ^*) of W by G. More fully, an associated prolongation of W by G with kernel K is a loop prolongation (L^*, ϕ^*) of W by G with $\phi^{*-1}(1) = K$, satisfying the usual conditions (5.4)-(5.8), and such that there exists a set of representatives $u^*(w) \in L^*$, with $\phi^*u^*(w) = w$, $u^*(1) = 1$ and

$$u^*(\bar{x}_1)u^*(\bar{x}_2) = u^*(\bar{x}_1\bar{x}_2)\,; \tag{8.1}$$

$$A^*(b, a, u^*(\bar{x}_2)) = 1 = A^*(a, u^*(\bar{x}_1), b)\,; \tag{8.2}$$

$$A^*(b, u^*(\bar{x}_1), u^*(\bar{x}_2)) = 1\,; \tag{8.3}$$

for all $a \in \phi^{*-1}(Y)$, $x_1, x_2 \in X$ and $b \in L^*$.

LEMMA 8d. *The correspondence $(L, \phi) \to (L^*, \phi^*)$ is one-one with the associated prolongations of W by G, to within an equivalence.*

I require to prove that the correspondence $(L, \phi) \to (L^*, \phi^*)$ has an inverse $(L^*, \phi^*) \to (L, \phi)$. Thus, let (L^*, ϕ^*) be an associated prolongation of W by G defined as above. Consider the set of elements $\phi^{*-1}(Y) \subset L^*$. They form a loop L_1. A homomorphism $\phi_1 : L_1 \to Y$ is defined by setting $\phi_1 = \phi^* | L_1$. The group X operates on L_1 through the representatives $u^*(\bar{x})$, i. e.:

$$x(a) = u^*(\bar{x}) \cdot a \qquad (x \in X\,;\, a \in L_1).$$

I prove that (L_1, ϕ_1) satisfies the conditions (7. 1)-(7. 7) above and thus that it is an X-prolongation of Y by G. Hence we have a correspondence $(L^*, \phi^*) \to (L_1, \phi_1)$ of associated prolongations of W by G to X-prolongations of Y by G.

Condition (7. 1) is obviously satisfied in virtue of the definition of (L_1, ϕ_1). Since $L_1 \supset K \supset G$, G is a subgroup of L_1. It is an operator subgroup, for

$$\begin{aligned} x(g) = u^*(\bar{x}) \cdot g &= \{\phi^* u^*(\bar{x})\} g \qquad \text{(by (5. 8))} \\ &= xg \end{aligned}$$

and by definition $xg \in G$.

The homomorphism ϕ_1 is an operator homomorphism, for if $x \in X$, $a \in L_1$, then

$$\begin{aligned} \{\phi_1(xa)\}\bar{x} = \phi_1\{(u^*(\bar{x}) \cdot a)u^*(\bar{x})\} &= \phi_1(u^*(\bar{x})a) \\ = \bar{x}\phi_1 a &= (x\phi_1 a)\bar{x}, \end{aligned}$$

so that $\phi_1(xa) = x\phi_1 a$. It is a trivial matter to prove that the kernel of ϕ_1 is K, which by definition contains G. This proves (7. 2).

The vector associator (A_0^3, A_1^3, A_2^3) of L_1 is defined as in section 6 above. By definition, if $a_1, a_2, a_3 \in L_1$, then

$$A_0^3(a_1, a_2, a_3) = A^*(a_1, a_2, a_3). \tag{8.4}$$

If in addition $x_1 \varepsilon X$, then

$$\begin{aligned}
\{x_1(a_2a_3)\}u^*(\bar{x}_1) &= \{u^*(\bar{x}_1)\cdot(a_2a_3)\}u^*(\bar{x}_1) = u^*(\bar{x}_1)(a_2a_3)\\
&= A^*(u^*(\bar{x}_1), a_2, a_3)\{(u^*(\bar{x}_1)a_2)a_3\}\\
&= A^*(u^*(\bar{x}_1), a_2, a_3)\{[(u^*(\bar{x}_1)\cdot a_2)u^*(\bar{x}_1)]a_3\}\\
&= A^*(u^*(\bar{x}_1), a_2, a_3)\{(x_1a_2)(u^*(\bar{x}_1)a_3)\} \qquad \text{(by (8. 2))}\\
&= A^*(u^*(\bar{x}_1), a_2, a_3)\{(x_1a_2)(u^*(\bar{x}_1)\cdot a_3)\}u^*(\bar{x}_1) \qquad \text{(by (8. 2))}\\
&= A^*(u^*(\bar{x}_1), a_2, a_3)\{(x_1a_2)(x_1a_3)\}u^*(\bar{x}_1).
\end{aligned}$$

Thus, by definition of $A_1{}^3$,

$$A_1{}^3(x_1; a_2, a_3) = A^*(u^*(\bar{x}_1), a_2, a_3). \tag{8. 5}$$

Similarly, if $x_1, x_2 \varepsilon X$,

$$\begin{aligned}
\{x_1(x_2a_3)\}u^*(\bar{x}_1\bar{x}_2) &= \{u^*(\bar{x}_1)\cdot(u^*(\bar{x}_2)\cdot a_3)\}\{u^*(\bar{x}_1)u^*(\bar{x}_2)\} \qquad \text{(by (8. 1))}\\
&= \{u^*(\bar{x}_1)(u^*(\bar{x}_2)\cdot a_3)\}u^*(\bar{x}_2) \qquad \text{(by (8. 3))}\\
&= u^*(\bar{x}_1)\{u^*(\bar{x}_2)a_3\} \qquad \text{(by (8. 2))}\\
&= A^*(u^*(\bar{x}_1), u^*(\bar{x}_2), a_3)\{[u^*(\bar{x}_1\bar{x}_2)\cdot a_3]u^*(\bar{x}_1\bar{x}_2)\}\\
&\qquad \text{(by (8. 1))}
\end{aligned}$$

Thus, by definition of $A_2{}^3$,

$$A_2{}^3(x_1, x_2; a_3) = A^*(u^*(\bar{x}_1), u^*(\bar{x}_2), a_3). \tag{8. 6}$$

The conditions (7. 3), (7. 4) and (7. 5) are now obviously satisfied in virtue of (8. 4), (8. 5) and (8. 6), and (5. 6) and (5. 7).

Finally, $a\cdot g = (\phi_1 a)g$ by definition of (L^*, ϕ^*) and

$$x(a\cdot g) = u^*(\bar{x})\cdot(a\cdot g).$$

But it is proved in [3] that in an ordinary loop prolongation (L, ϕ), defined as in section **5**, the elements of G associate with the elements of L and $a\cdot(b\cdot k) = A(a, b, k)[(ab)\cdot k]$ for all $a, b \varepsilon L$, $k \varepsilon K$. Thus, in particular $a\cdot(b\cdot g) = (ab)\cdot g$ for all $g \varepsilon G$. Hence, in (L^*, ϕ^*), we have

$$\begin{aligned}
x(a\cdot g) &= u^*(\bar{x})\cdot(a\cdot g) = (u^*(\bar{x})a)\cdot g\\
&= \{(u^*(\bar{x})\cdot a)u^*(\bar{x})\}\cdot g.
\end{aligned}$$

Thus again using the results of [3] quoted immediately above, we have

$$x(a\cdot g) = (u^*(\bar{x})\cdot a)(u^*(\bar{x})\cdot g) = (xa)\cdot(xg)$$

for all $x \varepsilon X$, $a \varepsilon L_1$, and $g \varepsilon G$. This proves (7. 6) and completes the proof that (L_1, ϕ_1) is an X-prolongation as required.

To complete the proof of the lemma, I require to show that the corre-

spondence $(L^*, \phi^*) \to (L_1, \phi_1)$ is the inverse of $(L, \phi) \to (L^*, \phi^*)$. Suppose therefore that given (L^*, ϕ^*), I form (L_1, ϕ_1) and then from (L_1, ϕ_1) I form the associated prolongation $(L_1{}^*, \phi_1{}^*)$ as above. Define a mapping $\theta\colon (L_1{}^*, \phi_1{}^*) \to (L^*, \phi^*)$ by setting $\theta(a, x) = au^*(\bar{x})$, for all $a \in L_1$, $x \in X$. I prove that θ is an equivalence of the prolongations as defined in section **5** above. It is certainly a homomorphism, for if $a_1, a_2 \in L_1$, $x_1, x_2 \in X$, then

$$\begin{aligned}\theta(a_1, x_1)\theta(a_2, x_2) &= (a_1u^*(\bar{x}_1))(a_2u^*(\bar{x}_2)) = \{a_1(u^*(\bar{x}_1)a_2)\}u^*(\bar{x}) \qquad \text{(by (8.2))}\\ &= \{a_1[(u^*(\bar{x}_1)\cdot a_2)u^*(\bar{x}_1)]\}u^*(\bar{x}_2) = \{a_1[(x_1a_2)u^*(\bar{x}_1)]\}u^*(\bar{x}_2)\\ &= [a_1(x_1a_2)]u^*(\bar{x}_1\bar{x}_2) \qquad \text{(by (8.1), (8.2) \& (8.3))}\\ &= \theta(a_1(x_1a_2), x_1x_2) = \theta\{(a_1, x_1)(a_2, x_2)\}.\end{aligned}$$

It is onto L^*, for if $a \in L^*$ and $\phi^*a^* = y\bar{x}$, we may solve the equation $a = zu^*(\bar{x})$ for a unique solution $z = a_1$, say, in which case $a = \theta(a_1, x)$. It is an isomorphism, for if $\theta(a, x) = 1$, then $au^*(\bar{x}) = 1$. Thus $\phi^*(au^*(\bar{x})) = (\phi^*a)\bar{x} = 1$ and $\bar{x} = 1$ and $u^*(\bar{x}) = 1$. Hence $a = 1$ and $(a, x) = 1$. Finally, $\phi^*\theta = \phi_1{}^*$, and θ is an equivalence. This completes the proof of the lemma.

9. The classification of the associated prolongations. In this section I use the results of [3] in order to classify the associated prolongations. First, using the representatives $u^*(w)$ of the associated prolongations (L^*, ϕ^*) defined above, we define the characteristic q-dimensional cochain $g_{u^{*q}}$ of (L^*, ϕ^*), which is in the group $C_*{}^q(W, Z) \subset C^q(W, Z)$, $q > 2$, where Z is the centre of K. We have, as in section **5** above,

$$g_{u^{*2}}(w_1, w_2)u^*(w_1w_2) = u^*(w_1)u^*(w_2),$$

$$g_{u^{*2n+1}}(w_1, \cdots, w_{2n}, w) = A^*(u^*(w_1), \cdots, u^*(w_{2n}), u^*(w)) \qquad (n \geq 1),$$

$$g_{u^{*2n+2}}(w_1, \cdots, w_{2n}, w, w') = A^*(u^*(w_1), \cdots, u^*(w_{2n}), g_{u^{*2}}(w, w')) \qquad (n \geq 1),$$

for all $w_1, \cdots, w_{2n}, w, w' \in W$. By definition of A^* and by (8.1)-(8.3), we have $g_{u^{*r}}(\bar{x}_1, \cdots, \bar{x}_q) = 1$ for all $x_1, \cdots, x_q \in X$. Thus $g_{u^{*q}} \in C_*{}^q(W, Z)$, for $r > 2$.

Thus we may quote the results of [3], i. e. Theorems EM1 and EM2 of section **5** above, with $H_*{}^n(W, G)$ for $H^n(W, G)$, to classify the associated prolongations, provided we can prove the existence

THEOREM 9a. *Given a cochain $g_*{}^q \in C_*{}^q(W, G)$ there is an associated*

prolongation (L^*, ϕ^*) *of* W *by* G *and a system of representatives* $u^*(w)$, *satisfying* (8.1)-(8.3) *above, such that*

$$g_{u^*}{}^q = g_*{}^q$$

and

$$(L^*, \phi^*) \in G_{n-1}{}^K \qquad (q = 2n, n > 0)$$

$$(L^*, \phi^*) \in G_n{}^L \qquad (q = 2n+1, n > 0),$$

where the classes $G_n{}^K$, $O_n{}^L$, *etc. are as defined in section* **5**. *If further* $g_{u^*}{}^q \in Z_*{}^q(W, G)$ *then also*

$$(L^*, \phi^*) \in O_n{}^L \cap G_{n-1}{}^K \qquad (q = 2n, n > 0),$$

$$(L^*, \phi^*) \in O_n{}^K \cap G_n{}^L \qquad (q = 2n+1, n > 0).$$

The prolongations (L^*, ϕ^*) *can be chosen so as to have an Abelian kernel.*

This is, of course, Theorem 12.1 of [3] with ordinary prolongations replaced by associated prolongations.

We first construct the kernel K of the desired prolongation. It is the direct sum $K_2 + K_3 + \cdots + K_{q-1} + G$, where K_m is a free Abelian group generated by the symbols $[w_1, y_2, w_3, w_4, \cdots, w_m]$ for all $w_i \in W$ and $y_2 \in Y \subset W$; and $K_m = G$ for $m \geq q$. Next we define cochains $h_*{}^m \in C_*{}^m(W, K)$, for $m \geq 2$ by setting

$$h_*{}^m(w_1, w_2, \cdots, w_m) = \begin{cases} [w_1, y_2, w_3, \cdots, w_m] & \text{if } w_i, y_2 \neq 1 \quad (1 \leq i \leq m) \\ 0 & \text{if some } w_i = 1 \quad (1 \leq i \leq m), \end{cases}$$

for $2 \leq m < q$, and $h_*{}^q = g_*{}^q$, $h_*{}^{q+1} = \delta_* g_*{}^q$, $h_*{}^m = 0$, $(m > q+1)$ $(w_i = y_i \bar{x}_i$ throughout). Further, for each $w_1 \in W$ we define an endomorphism $\alpha(w_1)$ of K by specifying:

$$\alpha(w_1) g = w_1 g$$

$$\alpha(w_1)[w_2, \cdots, w_{m+1}] = h_*{}^{m+1}(w_1, \cdots, w_{m+1})$$

$$- \sum_{i=1}^{m} (-1)^i h_*{}^m(w_1, \cdots, w_i w_{i+1}, \cdots, w_{m+1}) - (-1)^{m+1} h_*{}^m(w_1, \cdots, w_m),$$

for $2 \leq m < q$.

From here the proof is identical with that of Theorem 12.1 in [3]. I shall therefore not complete it, but go on to prove that the prolongation (L^*, ϕ^*) constructed from K by the methods of [3] is indeed an associated prolongation.

L^* is defined exactly as in [3] to be the set of all pairs (k, w), $k \in K$, $w \in W$ with $(k_1, w_1)(k_2, w_2) = (k_1[\alpha(w_1)k_2]h_*{}^2(w_1, w_2), w_1 w_2)$. The homo-

morphism $\phi^*: L^* \to W$ is defined by setting $\phi^*(k, w) = w$, and the representatives $u^*(w)$ are chosen by setting $u^*(w) = (1, w)$. If we now define the characteristic cochain $g_{u^*}{}^q$ of L^* in the usual way, we have $g_{u^*}{}^q = g_*{}^q$ (again exactly as in [3], Theorem 12.1). We require to show that (L^*, ϕ^*) satisfies (8.1)-(8.3). Certainly (8.1) holds, for if $w_1, w_2 \in W$, then, as in [3], $h_*{}^2(w_1, w_2)u^*(w_1w_2) = u^*(w_1)u^*(w_2)$. But $h_*{}^2(\bar{x}_1, \bar{x}_2) = 1$, for all $x_1, x_2 \in X$. Thus $u^*(\bar{x}_1\bar{x}_2) = u^*(\bar{x}_1)u^*(\bar{x}_2)$, which is (8.1). Also if $k_1, k_2 \in K$, $y_1 \in Y$, $w_2 \in W$, $x \in X$, then

$$\begin{aligned}\{(k_1, y_1)(1, \bar{x})\}(k_2, w_2) &= (k_1, y_1\bar{x})(k_2, w_2)\\ &= (k_1(\alpha(y_1\bar{x}))k_2)h_*{}^2(y_1\bar{x}, w_2), y_1\bar{x}w_2)\end{aligned}$$

and

$$\begin{aligned}(k_1, y_1)\{(1, \bar{x})(k_2, w_2)\} &= (k_1, y_1)((\alpha(\bar{x})k_2)h_*{}^2(\bar{x}, w_2), \bar{x}w_2)\\ &= (k_1\{\alpha(y_1)[(\alpha(\bar{x})k_2)h_*{}^2(\bar{x}, w_2)]\}h_*{}^2(y_1, \bar{x}w_2), y_1\bar{x}w_2).\end{aligned}$$

But $\alpha(y_1)\alpha(\bar{x}) = \alpha(y_1\bar{x})$, and by definition of $h_*{}^m$, $(\delta_*h_*{}^2)(y_1, \bar{x}, w_2) = 1$. Thus $A^*((k_1, y_1), (1, \bar{x}), (k_2, w_2)) = 1$, and since any element of $\phi^{*-1}(Y)$ is of the form (k_1, y_1), we have $A^*(a, u^*(\bar{x}), b) = 1$ for all $a \in \phi^{*-1}(Y)$, $x \in X$, and $b \in L^*$. This proves the second part of (8.2). Again, if $k_1, k_2 \in K$, $w_1 \in W$, $y_2 \in Y$, $x_2 \in X$,

$$\begin{aligned}\{(k_1, w_1)(k_2, y_2)\}(1, \bar{x}_2) &= (k_1(\alpha(w_1)k_2)h_*{}^2(w_1, y_2), w_1y_2)(1, \bar{x}_2)\\ &= (k_1(\alpha(w_1)k_2)h_*{}^2(w_1, y_2), w_1y_2\bar{x}_2);\end{aligned}$$

since $h_*{}^2(w_1y_2, \bar{x}_2) = 1$. Also, since $h_*{}^2(y_2, \bar{x}_2) = 1$,

$$\begin{aligned}(k_1, w_1)\{(k_2, y_2)(1, \bar{x}_2)\} &= (k_1, w_1)(k_2, y_2\bar{x}_2) = (k_1(\alpha(w_1)k_2)h_*{}^2(w_1, y_2\bar{x}_2), w_1y_2\bar{x}_2)\\ &= (k_1(\alpha(w_1)k_2)h_*{}^2(w_1, y_2), w_1y_2\bar{x}_2)\\ &= \{(k_1, w_1)(k_2, y_2)\}(1, \bar{x}_2),\end{aligned}$$

from above. Thus $A^*((k_1, w_1), (k_2, y_2), (1, \bar{x}_2)) = 1$, and since any element of $\phi^{*-1}(Y)$ if of the form (k_2, y_2) we have $A^*(b, a, u^*(\bar{x})) = 1$ for all $b \in L^*$. $a \in \phi^{*-1}(Y)$, and $x \in X$. This is the first part of (8.2).

Finally, if $k \in K$, $w \in W$, $x_1, x_2 \in X$,

$$(k, w)\{(1, \bar{x}_1)(1, \bar{x}_2)\} = (k, w)(1, \bar{x}_1\bar{x}_2) = (k, w\bar{x}_1\bar{x}_2)$$

and

$$\{(k, w)(1, \bar{x}_1)\}(1, \bar{x}_2) = (k, w_1\bar{x}_1)(1, \bar{x}_2) = (k, w\bar{x}_1\bar{x}_2),$$

so that $A^*(b, u^*(\bar{x}_1), u^*(\bar{x}_2)) = 1$, for all $b \in L^*$ and $x_1, x_2 \in X$, which proves (8.3).

This completes the proof of the modified existence theorem, as required for a quotation of Theorems EM1 and EM2. Thus, we have

THEOREM EM1*.

$$O_n{}^K \cap G_n{}^L/O_n{}^L \otimes G_{n-1}{}^K \approx H_*{}^{2n+1}(W, G)$$

where the classes $O_n{}^K$, $G_n{}^L$, etc. are classes of associated prolongations of W by G,

and a similar Theorem EM2* replaces Theorem EM2.

10. The classification of X-prolongations. In order to carry out the classification of the X-prolongations of Y by G, I require four types of X-prolongations (L, ϕ). They are $O_n{}^{*L}$, $O_n{}^{*K}$, $G_n{}^{*L}$ and $G_n{}^{*K}$. If $x_1, \cdots, x_{2n} \varepsilon X$, $a_1, \cdots, a_{2n+1} \varepsilon L$, $k \varepsilon K$, they are defined by the conditions

$$O_n{}^{*L}\colon\ A_r{}^{2n+1}(x_1, \cdots, x_r; a_{r+1}, \cdots, a_{2n}, a_{2n+1}) = 1,$$

$$O_n{}^{*K}\colon\ A_r{}^{2n+1}(x_1, \cdots, x_r; a_{r+1}, \cdots, a_{2n}, k) = 1,$$

$$G_n{}^{*L}\colon\ A_r{}^{2n+1}(x_1, \cdots, x_r; a_{r+1}, \cdots, a_{2n}, a_{2n+1}) \varepsilon G,$$

$$G_n{}^{*K}\colon\ A_r{}^{2n+1}(x_1, \cdots, x_r; a_{r+1}, \cdots, a_{2n}, k) \varepsilon G,$$

where in each case $r = 0, \cdots, 2n$.

I prove that if (L^*, ϕ^*) is the associated prolongation of (L, ϕ), then the one-one correspondence $(L, \phi) \to (L^*, \phi^*)$ preserves these classes. We have the

LEMMA 10a. *If (L, ϕ) belongs to one of the types $O_n{}^{*L}, \cdots, G_n{}^{*K}$ then (L^*, ϕ^*) belongs to the corresponding type $O_n{}^L, \cdots, G_n{}^K$, and vice versa, i. e. if $(L, \phi) \varepsilon O_n{}^{*L}$ then $(L^*, \phi^*) \varepsilon O_n{}^L$, etc.*

This is an immediate consequence of the

LEMMA 10b.

$$A_r{}^{2n+1}(x_1, \cdots, x_r; a_{r+1}, \cdots, a_{2n+1}) = A^*(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r; a_{r+1}, \cdots, a_{2n+1})$$

for all $x_i \varepsilon X$, $a_j \varepsilon L$ and $0 \leq r < 2n$.

We have already seen that this is true when $n = 1$, in Lemma 8d, with $\boldsymbol{x} = u^*(\bar{x})$. Suppose that the lemma is true for $n \leq k$, then by definition, if $r \geq 2$

$$A_r{}^{2k+3}(x_1, \cdots, x_r; a_{r+1}, \cdots, a_{2n+3})$$
$$= A_2{}^3(x_1, x_2; A_{r-2}{}^{2k+1}(x_3, \cdots, x_r; a_{r+1}, \cdots, a_{2n+3}))$$
$$= A^*(\boldsymbol{x}_1, \boldsymbol{x}_2, A^*(\boldsymbol{x}_3, \cdots, \boldsymbol{x}_r, a_{r+1}, \cdots, a_{2n+3}))$$
$$= A^*(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r, a_{r+1}, \cdots, a_{2n+3}).$$

Similar inductive arguments prove the lemma then $r = 0, 1$.

In addition to the classes $O_n{}^{*L}, \cdots, G_n{}^{*K}$, we require a fifth class, $R_*{}^n$, of *n-reducible X-prolongations*. These are X-prolongations which contain (in the obvious sense) an X-prolongation (L', ϕ') of the class $O_n{}^{*K}$. If (L^*, ϕ^*) is the associated prolongation of (L, ϕ) and (L, ϕ) contains $(L', \phi') \in O_n{}^{*K}$, then from its definition it follows that (L^*, ϕ^*) contains the prolongation associated with (L', ϕ'). Moreover it will follow that this associated prolongation is in the class $O_n{}^K$. Conversely, given (L^*, ϕ^*) containing $(L'^*, \phi'^*) \in O_n{}^K$, it is easily seen that the X-prolongation (L, ϕ), constructed from (L^*, ϕ^*) as in section **8** above, contains an X-prolongation $(L', \phi') \in O_n{}^{*K}$ whose associated prolongation is (L'^*, ϕ'^*). Thus the one-one correspondence $(L, \phi) \leftrightarrow (L^*, \phi^*)$ carries $R_n{}^*$ onto R_n.

Finally we require the products of the classes $O_n{}^{*L}$, etc. The product $A \otimes B$ of two classes A, B is defined to be the set $\{L_1 \otimes L_2\}$ of X-prolongations of Y by G, for all $L_1 \in A$, $L_2 \in B$, where $L_1 \otimes L_2$ is defined as in section **7** above. It is a trivial matter to prove that the prolongation associated with a product of X-prolongations is the product of the associated prolongations. Thus products of the classes $O_n{}^{*L}, \cdots, G_n{}^{*K}$ are preserved. Hence using Theorems EM1* and EM2*, together with the fact that

$$H_*{}^n(W, G) \approx \mathfrak{H}_{n-1}{}^n(X, Y; G),$$

we have the

THEOREM A.

$$O_n{}^{*K} \cap G_n{}^{*L}/O_n{}^{*L} \otimes G_{n-1}{}^{*K} \approx \mathfrak{H}_{2n}{}^{2n+1}(X, Y; G) \qquad (n > 0)$$

and the

THEOREM B.

$$O_{n+1}{}^{*L} \cap G_n{}^{*K}/R_n{}^* \otimes G_n{}^{*L} \approx \mathfrak{H}_{2n+1}{}^{2n+2}(X, Y; G) \qquad (n > 0)$$

where in each case the system on the left is a group.

11. The direct approach. Theorems A and B can be obtained directly by methods entirely analogous to those of [3], by using vector characteristic cochains $(c_0, \cdots, c_{q-1})^q \in \mathfrak{C}_{q-1}{}^q(X, Y; K)$, which are associated with an X-prolongation (L, ϕ). These cochains are defined by

$$c_q(x_1, \cdots, x_q; y_{q+1}, \cdots, y_{2n}, y) = A_q^{2n+1}(x_1, \cdots, x_q; u(y_{q+1}), \cdots, u(y)) \qquad (n \geq 1)$$

$$c_q(x_1, \cdots, x_q; y_{q+1}, \cdots, y_{2n}, y, y') = A_q^{2n+2}(x_1, \cdots, x_q; u(y_1), \cdots, u(y_{2n}), k_0(y, y')) \qquad (n \geq 1)$$

$$c_{2n+1}(x_1, \cdots, x_{2n+1}, y) = A_{2n}^{2n+2}(x_1, \cdots, x_{2n}; k_1(x_{2n+1}, y))$$

where $q = 0, \cdots, 2n$, $\phi u(y_i) = y_i$, $u(1) = 1$, $x_j \in X$, $y_i \in Y$, for all i, j as required, and

$$k_0(y, y')u(yy') = u(y)u(y'), \qquad k_1(x_{2n+1}; y)u(x_{2n+1}y) = x_{2n+1}u(y).$$

These cochains $(c_0, \cdots, c_{q-1})^q$ are then homomorphisms which induce the isomorphism of Theorems A and B in the same way as the characteristic cochains $g_u{}^q$ induce the isomorphisms of Theorems EM1 and EM2.

KING'S COLLEGE, ABERDEEN, SCOTLAND.

REFERENCES.

[1] R. Baer, "Automorphismen von Erweiterungsgruppen," *Actualités scientifique et Industrielles*, No. 205, Paris, 1934.

[2] S. Eilenberg, "Topological methods in abstract algebra. Cohomology theory of groups," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 3-37.

[3] S. Eilenberg and S. Maclane, "Algebraic cohomology groups and loops," *Duke Mathematical Journal*, vol. 14 (1947), pp. 435-463.

[4] J. H. C. Whitehead, "On group extensions with operators," *Quarterly Journal of Mathematics (Oxford)*, vol. 3 (1950), pp. 219-228.

DOUBLY STOCHASTIC MATRICES AND THE DIAGONAL OF A ROTATION MATRIX.* [1]

By ALFRED HORN.

An elegant theorem of Hardy, Littlewood and Polya (Theorem 1 below) has been generalized and applied in various ways in recent literature. In this paper we shall present an extension of their theorem (Theorm 4). As an application we prove that the set of all diagonals of rotation matrices of order n is equal to the convex hull of those points $(\pm 1, \cdots, \pm 1)$ of which an even number of coordinates are -1. From this, we determine the set of all diagonals of orthogonal or of unitary matrices. It would be interesting to prove without Theorem 4 that the set of diagonals of rotation matrices is convex.

1. Definitions and notation. If A is a matrix, then A_{ij} will denote its i, j component and we write $A = (A_{ij})$. A similar convention will be used for vectors. An n-vector is a vector with n components. A unit vector is a vector of length one. The diagonal of a matrix A is the vector $(A_{11}, \cdots, A_{nn})$. A diagonal matrix is one whose non-diagonal elements vanish. The identity matrix is the diagonal matrix with diagonal $(1, \cdots, 1)$.

A permutation matrix is a matrix with elements all 0 or 1 such that every row and every column contains exactly one element equal to 1. A doubly stochastic (d. s.) matrix is a matrix P such that $P_{ij} \geqq 0$, $\sum_i P_{ij} = \sum_j P_{ij} = 1$ for all i and j. We use the symbols A', $\bar{A}$ and A^* for transpose, conjugate and conjugate-transpose respectively. By an orthogonal matrix, we mean a real matrix A such that $A' = A^{-1}$, A rotation matrix, or rotation, is an orthogonal matrix with determinant $+1$. A unitary matrix is a matrix A such that $A^* = A^{-1}$.

If x is a real n-vector, then $H(x)$ is the convex hull of all the points $(x_{\alpha_1}, \cdots, x_{\alpha_n})$, α varying over all permutations of $(1, \cdots, n)$. By $S^n{}_k$, we mean the set of all k termed sequences σ of integers for which $1 \leqq \sigma_1 < \cdots < \sigma_k \leqq n$. The only member of $S^n{}_0$ is the empty sequence.

* Received April 16, 1953.

[1] This paper was written while the author received partial support from the Office of Naval Research.

If x is a complex number, $\Re x$ denotes its real part. Occasionally we shall use i both as a summation index and to denote the imaginary unit $(-1)^{\frac{1}{2}}$. But we shall do this in such a way as to avoid confusion.

2. Doubly stochastic and ortho-stochastic matrices.

THEOREM 1. *Let x, y be real n-vectors. Then the following statements are equivalent:*

(1a) $y = Px$, *where P is a d. s. matrix.*

(1b) $y \in H(x)$.

(1c) $\max_{\sigma \in S^n_k} \sum_{i=1}^{k} y_{\sigma_i} \leqq \max_{\sigma \in S^n_k} \sum_{i=1}^{k} x_{\sigma_i}, \quad 1 \leqq k \leqq n$ and $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i$.

(1d) $\sum_{i=1}^{n} f(y_i) \leqq \sum_{i=1}^{n} f(x_i)$ *for any convex function f whose domain contains all the numbers x_i, y_i, $1 \leqq i \leqq n$.*

Proof. The equivalence of (1a), (1c) and (1d) is proved in [1], pp. 49 and 89. Also (1b) is equivalent to (1a) by virtue of the following theorem of G. Birkhoff [2].

THEOREM 2. *A matrix is d. s. if and only if it lies in the convex hull of the set of all permutation matrices.*

In case $x_1 \geqq \cdots \geqq x_n$, the conditions (1c) are equivalent to

(1e) $$\sum_{i=n-k+1}^{n} x_i \leqq \sum_{i=1}^{k} y_{\sigma_i} \leqq \sum_{i=1}^{k} x_i, \quad \sigma \in S^n_k, \quad 1 \leqq k \leqq n.$$

We take this opportunity to point out a companion to Theorem 1, various parts of which are scattered through the literature.

THEOREM 3. *Let x, y be n-vectors for which $x_i \geqq 0$, $y_i \geqq 0$, $1 \leqq i \leqq n$. Then the following statements are equivalent:*

(2a) $y = Px$, *where P is a matrix such that $P_{ij} \geqq 0$, $\sum_{i=1}^{n} P_{ij} \leqq 1$, $\sum_{j=1}^{n} P_{ij} \leqq 1$, $1 \leqq i \leqq n$, $1 \leqq j \leqq n$.*

(2b) $\max_{\sigma \in S^n_k} \sum_{i=1}^{k} y_{\sigma_i} \leqq \max_{\sigma \in S^n_k} \sum_{i=1}^{k} x_{\sigma_i}, \quad 1 \leqq k \leqq n$.

(2c) $\sum_{i=1}^{n} f(y_i) \leqq \sum_{i=1}^{n} f(x_i)$ *for any convex non-decreasing function f whose domain contains all the numbers x_i, y_i, $1 \leqq i \leqq n$.*

Proof. The proof given in [1] for the implication (1d) → (1c) also proves (2c) → (2b). The fact that (2b) → (2c) is proved in [3]. Ky Fan [4] proves (2a) → (2b). He also states (2b) → (2a) and quotes a lemma used in his proof. Using the fact that (1c) → (1a), a simple proof may be obtained. It is also possible to derive (2a) → (2b) from (1a) → (1b) by using the fact that any matrix of the kind described in (2a) may be augmented to a d. s. matrix (of order $\leqq 2n$).

A special class of d. s. matrices consists of those matrices Q such that $Q_{ij} = U_{ij}^2$ where U is an orthogonal matrix. Let us call such matrices orthostochastic (o. s.) matrices. Not every d. s. matrix is an o. s. matrix. In fact it is not hard to see that the matrix

$$\tfrac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

cannot even be expressed as a product of o. s. matrices. (This example is due to A. J. Hoffman.) Nevertheless we have the following theorem.

THEOREM 4. *If x, y are n-vectors satisfying* (1c), *then there exists an o. s. matrix Q such that $y = Qx$.*

The proof of Theorem 4 will be simplified if we first prove a lemma.

DEFINITON. If $x_1 \geqq \cdots \geqq x_n$, and $1 \leqq m \leqq n$, let $T^m(x_1, \cdots, x_n)$ be the set of points $(z_1, \cdots, z_m)$ such that $\sum_{i=n-k+1}^{n} x_i \leqq \sum_{i=1}^{k} z_{\sigma_i} \leqq \sum_{i=1}^{k} x_i$ whenever $\sigma \in S^n{}_k$, $1 \leqq k \leqq m$. Also let $T^0(x_1, \cdots, x_n)$ be the empty set.

LEMMA. *If $x_1 \geqq \cdots \geqq x_n$ and $(y_1, \cdots, y_{n-1}) \in T^{n-1}(x_1, \cdots, x_n)$ and $n \geqq 2$, then there exists a set of $n-1$ real ortho-normal n-vectors $u^1, \cdots, u^{n-1}$ such that*

$$y_j = \sum_{i=1}^{n} (u^j{}_i)^2 x_i, \quad 1 \leqq j \leqq n-1. \tag{3}$$

Proof. The statement is obvious for $n = 2$. We proceed by induction. Suppose the lemma is true for $2 \leqq n < N$. Clearly $T^{N-1}(x_1, \cdots, x_N)$ consists of the points $(y_1, \cdots, y_{N-1})$ such that $(y_1, \cdots, y_{N-2}) \in T^{N-2}(x_1, \cdots, x_N)$ and such that $b \leqq y_{N-1} \leqq a$, where b is the largest of the numbers $b_\sigma = \sum_{i=N-k}^{N} x_i - \sum_{i=1}^{k} y_{\sigma_i}$, and a is the smallest of the numbers $a_\sigma = \sum_{i=1}^{k+1} x_i - \sum_{i=1}^{k} y_{\sigma_i}$ as σ ranges over $S^{N-2}{}_k$ and k varies from 0 to $N-2$. (When σ is the empty sequence, $b_\sigma = x_N$ and $a_\sigma = x_1$).

For each $\sigma \in S^{N-2}{}_k$, $0 \leqq k \leqq N-2$, let T_σ be the subset of $T^{N-2}(x_1, \cdots, x_N)$ on which $a = a_\sigma$, and let T'_σ be the subset on which $b = b_\sigma$. It is not hard to verify that T_σ consists of the points $(y_1, \cdots, y_{N-2})$ such that

$$(y_{\sigma_1}, \cdots, y_{\sigma_k}) \in T^k(x_1, \cdots, x_{k+1}) \text{ and } (y_{\sigma'_1}, \cdots, y_{\sigma'_{N-2-k}}) \in$$

$T^{N-2-k}(x_{k+2}, \cdots, x_N)$, where σ' is the sequence consisting of those integers from 1 to $N-2$ which are not among the terms of σ. Verify also that $T_\sigma = T'_{\sigma'}$. It follows that $T^{N-1}(x_1, \cdots, x_N) = \bigcup_{k=0}^{N-2} \bigcup_{\sigma \in S^{N-2}{}_k} S_\sigma$, where S_σ is the set of points $(y_1, \cdots, y_{N-1})$ such that $(y_1, \cdots, y_{N-2}) \in T_\sigma$ and $b_{\sigma'} \leqq y_{N-1} \leqq a_\sigma$.

Now let $(y_1, \cdots, y_{N-1})$ be a fixed point of $T^{N-1}(x_1, \cdots, x_N)$. Then $(y_1, \cdots, y_{N-1}) \in S_\sigma$ for some fixed $\sigma \in S^n{}_k$ and some fixed $k \leqq N-2$. We must find an ortho-normal sequence $u^1, \cdots, u^{N-1}$ for which (3) holds. By the induction hypothesis, there exist k real ortho-normal $k+1$-vectors $v^{\sigma_1}, \cdots, v^{\sigma_k}$ and $N-2-k$ real ortho-normal $N-1-k$-vectors $v^{\sigma'_1}, \cdots, v^{\sigma'_{N-2-k}}$ such that $y_{\sigma_m} = \sum_{i=1}^{k+1} (v^{\sigma_m}{}_i)^2 x_i$, $1 \leqq m \leqq k$, $y_{\sigma'_m} = \sum_{i=1}^{N-1-k} (v^{\sigma'_m}{}_i)^2 x_{k+1+i}$, $1 \leqq m \leqq N-2-k$. We define $N-2$ ortho-normal N-vectors $u^1, \cdots, u^{N-2}$ as follows: $u^{\sigma_m} = (v^{\sigma_m}{}_1, \cdots, v^{\sigma_m}{}_{k+1}, 0, \cdots, 0)$, $1 \leqq m \leqq k$; $u^{\sigma'_m} = (0, \cdots, 0, v^{\sigma'_m}{}_1, \cdots, v^{\sigma_m}{}'_{N-1-k})$, $1 \leqq m \leqq N-2-k$. We have $y_j = \sum_{i=1}^{N} (u^j{}_i)^2 x_i$, $1 \leqq j \leqq N-2$. If u ranges over the circle consisting of all real N-vectors orthogonal to $u^1, \cdots, u^{N-2}$, then the values of the sum

$$\sum_{i=1}^{N} (u_i)^2 x_k \tag{4}$$

fill out some interval. If we take $u = (v_1, \cdots, v_{k+1}, 0, \cdots, 0)$, where $(v_1, \cdots, v_{k+1})$ is a real unit $k+1$-vector orthogonal to $v^{\sigma_1}, \cdots, v^{\sigma_k}$, then (4) takes on the value $\sum_{i=1}^{k+1} x_i - \sum_{j=1}^{k} y_{\sigma_j} = a_\sigma$. Also if we take $u = (0, \cdots, 0, v_1, \cdots, v_{N-1-k})$, where $(v_1, \cdots, v_{N-1-k})$ is a real unit vector orthogonal to $v^{\sigma'_1}, \cdots, v^{\sigma'_{N-2-k}}$, then (4) takes on the value $\sum_{i=k+2}^{N} x_i - \sum_{j=1}^{N-2-k} y_{\sigma'_j} = b_{\sigma'}$. Therefore if $b_{\sigma'} \leqq y_{N-1} \leqq a_\sigma$, there exists a real unit N-vector u^{N-1} orthogonal to $u^1, \cdots, u^{N-2}$ such that $y_{N-1} = \sum_{i=1}^{N} (u^{N-1}{}_i)^2 x_i$.

Proof of Theorem 4. There is no loss of generality if we assume $x_1 \geqq \cdots \geqq x_n$. As mentioned above, condition (1c) is equivalent with (1e).

Furthermore it is easily verified that (1e) is equivalent to the conditions $(y_1, \cdots, y_{n-1}) \in T^{n-1}(x_1, \cdots, x_n)$, and $y_n = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n-1} y_i$. By the preceding lemma, there exists an ortho-normal system $u^1, \cdots, u^{n-1}$ such that (3) holds. If u^n is a real unit vector orthogonal to $u^1, \cdots, u^{n-1}$, it follows that $\sum_{i=1}^{n} (u^n_i)^2 x_i = \sum_{i=1}^{n} x_i - \sum_{j=1}^{n-1} y_j = y_n$. Thus if we set $Q_{ij} = (u^i_i)^2$, then $Q = (Q_{ij})$ is the desired o. s. matrix.

An immediate consequence of Theorem 4 is the following.

THEOREM 5. *A vector y can be the diagonal of a Hermitian matrix with eigenvalues $x_1, \cdots, x_n$ if and only if* (1c) *holds.*

Proof. A Hermitian matrix A has eigenvalues $x_1, \cdots, x_n$ if and only if $A = UBU^*$, where B is the diagonal matrix with diagonal $(x_1, \cdots, x_n)$, and U is a unitary matrix. Therefore y is the diagonal of such a matrix A if and only if $y_i = \sum_{j=1}^{n} |U_{ij}|^2 x_j$. Theorems 1 and 4 now yield the conclusion.

The neecssity of (1c) in Theorem 5 was pointed out by Schur [5], who showed that this leads to a simple proof of Hadamard's inequality for determinants.

It is interesting to compare Theorem 5 with the following theorem, proved in [6].

THEOREM 6. *Let $x_i > 0$, $1 \leqq i \leqq n$. Then there exists a Hermitian matrix A with eigenvalues $x_1, \cdots, x_n$ such that $\prod_{i=1}^{k} y_i = M_k$, $1 \leqq k \leqq n$, where M_k is the determinant formed from the first k rows and columns of A, if and only if $\max_{\sigma \in S^n_k} \prod_{i=1}^{k} y_{\sigma_i} \leqq \max_{\sigma \in S^n_k} \prod_{i=1}^{k} x_{\sigma_i}$ and $\prod_{i=1}^{n} y_i = \prod_{i=1}^{n} x_i$.*

3. Remarks on the complex case. If $y = Px$, where x, y are *complex* n-vectors, and P is a d. s. matrix, then there is an analogue to the implication (1a) $\rightarrow$ (1e). We can show that for each $\sigma \in S^n_k$, $\sum_{i=1}^{k} y_{\sigma_i}$ lies in the convex hull of all the points $\sum_{i=1}^{k} x_{\tau_i}$, $\tau \in S^n_k$. More generally, we have:

THEOREM 7. *If $y = Px$, where x, y are complex n-vectors, and P is a d. s. matrix, and if $c_1, \cdots, c_n$ are any complex numbers, then $\sum_{i=1}^{n} c_i y_i$ lies in*

the convex hull of all the points $\sum_{i=1}^{n} c_i x_{\alpha_i}$, $\alpha \in R^n$, *where* R^n *is the set of all permutations of* $(1, \cdots, n)$.

Proof. By Theorem 2, we may write $P = \sum_{\alpha \in R^n} a_\alpha P^\alpha$, where $P^\alpha{}_{ij} = 1$ when $\alpha_i = j$, and $P^\alpha{}_{ij} = 0$ otherwise, and $a_\alpha \geqq 0$, $\sum_\alpha a_\alpha = 1$. Therefore

$$\sum_{i=1}^{n} c_i y_i = \sum_{i,j=1}^{n} P_{ij} c_i x_j = \sum_\alpha a_\alpha \sum_{i,j=1}^{n} P^\alpha{}_{ij} c_i x_j$$

$$= \sum_\alpha a_\alpha \sum_{i=1}^{n} c_i \sum_{j=1}^{n} P^\alpha{}_{ij} x_j = \sum_\alpha a_\alpha \sum_{i=1}^{n} c_i x_{\alpha_i}.$$

We do not know whether the converse of Theorem 7 is true. However, there cannot be any simple analogue of (1e) → (1a). In fact there exists a pair x, y of 4-vectors such that $\sum_{i=1}^{4} c_i y_i$ lies in the convex hull of the points $\sum_{i=1}^{4} c_i x_{\alpha_i}$ for all *real* $c_1 \cdots, c_4$ and yet there exists no d. s. matrix P such that $y = Px$. For example take $y = (1/4 + 2i/3, -1/4 + 2i/3, 0, 0)$ and $x = (1, -1, i, i/3)$. Theorem 4 also breaks down in the complex case. In fact if $y = (0, 1/2 + i/2, -1/2 + i/2)$, $x = (1, -1, i)$, then there is one and only one d. s. matrix P for which $y = Px$, namely

$$P = \tfrac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

However the equivalence (1d) ↔ (1a) still holds for complex vectors [7].

4. The diagonal of a rotation.

THEOREM 8. *A vector* $(d_1, \cdots, d_n)$ *is the diagonal of a rotation of order n if and only if it lies in the convex hull of those points* $(\pm 1, \cdots, \pm 1)$ *of which an even number (possibly* 0*) of coordinates are* -1.

Proof. If R is a rotation, then its eigenvalues are all of modulus 1. The complex eigenvalues occur in conjugate pairs and an even number of the real eigenvalues are -1. Therefore we may say that if n is odd, say $n = 2m + 1$, then the eigenvalues are of the form $1, \exp(\pm i\theta_1), \cdots, \exp(\pm i\theta_m)$, where $0 \leqq \theta_j < 2\pi$. If $n = 2m$, the eigenvalues have the form $\exp(\pm i\theta_1), \cdots, \exp(\pm i\theta_m)$. In what follows, the case n even is practically identical with the case n odd. Accordingly we will only treat the case $n = 2m + 1$. It is well known that there exists an orthogonal matrix U such that $R = U'AU$, where

$$A = \begin{bmatrix} 1 & & & 0 \\ & B_1 & & \\ & & \ddots & \\ 0 & & & B_m \end{bmatrix}, \quad \text{and} \quad B_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}, \quad 1 \leqq i \leqq m.$$

Therefore

$$R_{jj} = U_{j1}^2 + U_{j2}^2 \cos\theta_1 + U_{j3}^2 \cos\theta_1 + \cdots + U_{j2m}^2 \cos\theta_m + U_{j2m+1}^2 \cos\theta_m, \qquad 1 \leqq j \leqq n.$$

By Theorem 4, it follows that $(d_1, \cdots, d_n)$ is the diagonal of a rotation with eigenvalues 1, $\exp(\pm i\theta_1), \cdots, \exp(\pm i\theta_m)$ if and only if it lies in $H(1, \cos\theta_1, \cos\theta_1, \cdots, \cos\theta_m, \cos\theta_m)$. Since $H(x)$ is unchanged if we permute the coordinates of x, we find that the set S of all diagonals of n-th order rotations has the form

$$S = \bigcup_{1 \geqq a_1 \geqq \cdots \geqq a_m \geqq -1} H(1, a_1, a_1, \cdots, a_m, a_m). \tag{5}$$

Let T be the convex hull of those points $(\pm 1, \cdots, \pm 1)$ of which an even number of coordinates are -1. To prove $S \subset T$, we need only show that if $1 \geqq a_1 \geqq \cdots \geqq a_m \geqq -1$, then $(1, a_1, a_1, \cdots, a_m, a_m) \in T$. This follows immediately from the formula

$$\begin{aligned} &\sum_{x_i = \pm 1,\, 1 \leqq i \leqq m} \prod_{j=1}^{m} \tfrac{1}{2}(1 + x_j a_j)(1, x_1, x_1, \cdots, x_m, x_m) \\ &= \sum_{x_i = \pm 1,\, 1 \leqq i \leqq m-1} \prod_{j=1}^{m-1} \tfrac{1}{2}(1 + x_j a_j)(1, x_1, x_1, \cdots, x_{m-1}, x_{m-1}, a_m, a_m) \\ &= \cdots = (1, a_1, a_1, \cdots, a_m, a_m). \end{aligned}$$

To show that $T \subset S$, we first note that each of the vertices of T is obviously in S, by (5). Therefore to complete the proof, we need only show that S is convex. Suppose x, y are distinct points of S, and $0 < \lambda < 1$. By (5), there exist sequences $\{a_i\}$, $\{b_i\}$ such that $1 \geqq a_1 \geqq \cdots \geqq a_m \geqq -1$, $1 \geqq b_1 \geqq \cdots \geqq b_m \geqq -1$,

$$x \in H(1, a_1, a_1, \cdots, a_m, a_m), \quad \text{and} \quad y \in H(1, b_1, b_1, \cdots, b_m, b_m).$$

By using the equivalence of (1b) and (1c) it is easy to show that

$$\lambda x + (1-\lambda) y \in H(1, \lambda a_1 + (1-\lambda) b_1, \lambda a_1 + (1-\lambda) b_1, \cdots, \lambda a_m + (1-\lambda) b_m, \lambda a_m + (1-\lambda) b_m) \subset S.$$

THEOREM 9. *Let x be a real n-vector, $n \geqq 2$. The following statements are equivalent*

(6a) $\sum_{j \neq i} x_j \leqq n - 2 + x_i$, *and* $0 \leqq x_i \leqq 1, 1 \leqq i \leqq n$.

(6b) *x lies in the convex hull of B, where B is the set of those points with coordinates all 0 or 1 for which the number of 0 coordinates is different from one.*

(6c) *x is the diagonal of a rotation, and $x_i \geqq 0$, $1 \leqq i \leqq n$.*

(6d) *x is the diagonal of a d. s. matrix.*

Proof. Given any set A of k integers between 1 and n, there exists a permutation which leaves invariant the members of A, and only these, as long as $k \neq n - 1$. Therefore the set of diagonals of permutation matrices of order n is exactly B. Theorem 2 now shows that (6d) and (6b) are equivalent. We conclude the proof by deriving the implications (6b) $\rightarrow$ (6c) $\rightarrow$ (6a) $\rightarrow$ (6d).

(6b) $\rightarrow$ (6c): Since the set S of diagonals of rotations is convex, we need only prove $B \subset S$. If $x \in B$, then some permutation of x has the form $(y_1, \cdots, y_n)$, in which $y_i = 0$, $1 \leqq i \leqq k$, $y_i = 1$, $k + 1 \leqq i \leqq n$, and $k \neq 1$. If k is even, then $y \in S$ by (5). (When $n = 2m$, (5) remains valid if we drop the 1 in $H(1, a_1, a_1, \cdots, a_m, a_m)$). If k is odd, then $k \geqq 3$, and using the equivalence of (1b) and (1c), it is easy to prove that $y \in H(z)$, where $z_i = 1$, $1 \leqq i \leqq n - k + 1$, $z_{n-k+2} = \frac{1}{2}$, $z_{n-k+3} = -\frac{1}{2}$, and $z_i = 0$ otherwise. Again $H(z) \subset S$ by (5).

(6c) $\rightarrow$ (6a): Let x be the diagonal of a rotation of order n. If $n = 2m + 1$, then by (5), $x \in H(1, a_1, a_1, \cdots, a_m, a_m)$ for some sequence $\{a_i\}$ with $1 \geqq a_1 \geqq \cdots \geqq a_m \geqq -1$. Therefore by (1e),

$$\sum_{j \neq i} x_j \leqq 1 + 2a_1 + \cdots + 2a_{m-1} + a_m \leqq n - 2 + x_i.$$

The case n even is analogous.

(6a) $\rightarrow$ (6d): Suppose x satisfies (6a). We may exclude the obvious case $x_i = 1$ for all i. Let $y_i = 1 - x_i$. Then $0 \leqq y_i \leqq 1$ and $2y_i \leqq \sum_{j=1}^{n} y_j \neq 0$. Let $\beta = \sum_{i=1}^{n} y_i$ and set $z_i = 2y_i/\beta$. Then $0 \leqq z_i \leqq 1$ and $\sum_{i=1}^{n} z_i = 2$. By the equivalence of (1b) and (1c), we have $(z_1, \cdots, z_n) \in H(1, 1, 0, \cdots, 0)$.

Therefore $(z_1, \cdots, z_n) = \sum_{1 \leqq i < j \leqq n} \alpha_{ij} u^{ij}$, where u^{ij} is the n-vector whose i-th and j-th components are 1, and whose other components are 0, and $\alpha_{ij} \geqq 0$, $\sum_{i<j} \alpha_{ij} = 1$. Therefore $(1 - x_1, \cdots, 1 - x_n) = \sum_{i<j} \gamma_{ij} u^{ij}$, where $\gamma_{ij} = \beta \cdot \alpha_{ij}/2$. If we define $P_{ii} = x_i$ and $P_{ij} = P_{ji} = \gamma_{ij}$ for $1 \leqq i \leqq j \leqq n$, we obtain a symmetric d. s. matrix $P = (P_{ij})$ with the desired property.

THEOREM 10. *For $n \geqq 2$, a real vector $(d_1, \cdots, d_n)$ is the diagonal of an orthogonal matrix if and only if the vector $x = (|d_1|, \cdots, |d_n|)$ satisfies any one of the conditions of Theorem* 9.

Proof. Suppose x satisfies (6c). If x is the diagonal of a rotation, then d is the diagonal of an orthogonal matrix, since an orthogonal matrix remains orthogonal when we multiply any row by -1.

Now suppose d is the diagonal of an orthogonal matrix. Then at least one of the two vectors $(|d_1|, \cdots, |d_{i-1}|, \pm d_i, |d_{i+1}|, \cdots, |d_n|)$ is the diagonal of a rotation. The proof of (6c) $\rightarrow$ (6a) did not use the hypothesis $x_i \geqq 0$. Therefore we have $\sum_{j \neq i} |d_j| \leqq n - 2 \pm d_i \leqq n - 2 + |d_i|$, which proves (6a).

THEOREM 11. *For $n \geqq 2$, a complex vector $(d_1, \cdots, d_n)$ is the diagonal of a unitary matrix if and only if the vector $x = (|d_1|, \cdots, |d_n|)$ satisfies any one of the conditions of Theorem* 9.

Proof. Suppose x satisfies (6c). Then x is the diagonal of a rotation R. If we multiply the i-th row of R by $\arg d_i$, whenever $d_i \neq 0$, then R becomes a unitary matrix with diagonal $(d_1, \cdots, d_n)$.

To prove the converse (the idea of this proof is due to J. von Neumann), let $\alpha = \sup (|d_1| + \cdots + |d_{n-1}| - |d_n|)$ as d varies over all diagonals of unitary matrices of order n. In order to prove (6a), we need only show $\alpha = n - 2$. Clearly $\alpha \geqq n - 2$, as is seen by considering the identity matrix. Now let U be a unitary matrix for which $|U_{11}| + \cdots + |U_{n-1\,n-1}| - |U_{nn}| = \alpha$. By multiplying each row by a proper factor, we may assume that $U_{ii} \geqq 0$, $1 \leqq i \leqq n$. By the maximal property of U, we have $U_{ii} \geqq U_{nn}$, since otherwise an interchange of rows of U would increase the value of $\sum_{i=1}^{n-1} U_{ii} - U_{nn}$. Thus it involves no loss of generality to assume $U_{11} \geqq \cdots \geqq U_{nn} \geqq 0$.

If $U_{n-1\,n-1} = 0$, then $\alpha = \sum_{i=1}^{n-2} U_{ii} \leqq n - 2$ and (6a) is proved. Henceforth assume $U_{n-1\,n-1} > 0$. For $1 \leqq p < q \leqq n$, let V^{pq} be the matrix of

order n for which $V^{pq}{}_{pp} = V^{pq}{}_{qq} = \cos\theta$, $V^{pq}{}_{pq} = -\bar{V}^{pq}{}_{qp} = \sin\theta \exp(i\beta)$, where θ, β are arbitrary real numbers, and $V^{pq}{}_{ij} = \delta_{ij}$ for $i \neq p, q$, and $j \neq p, q$. If $W = UV^{12}$, we have $\sum_{i=1}^{n-1} |W_{ii}| - |W_{nn}| \leqq \sum_{i=1}^{n-1} U_{ii} - U_{nn}$. Therefore

$$(7) \quad |U_{11}\cos\theta - U_{12}\sin\theta\exp(-i\beta)| + |U_{21}\sin\theta\exp(i\beta) + U_{22}\cos\theta| \leqq U_{11} + U_{22}.$$

The expansion of the left side of (7) in powers of θ begins with

$$(U_{11} + U_{22}) + \theta[\Re(\exp(i\beta)(U_{21} - \bar{U}_{12}))] + \cdots.$$

Consequently the coefficient of θ must vanish for all β. This can occur only if $U_{21} = \bar{U}_{12}$. Similarly $U_{ji} = \bar{U}_{ij}$ for $1 \leqq i \leqq j \leqq n-1$. From this it follows that $|U_{in}| = |U_{ni}|$, $1 \leqq i \leqq n$.

If $U_{nn} > 0$, then by multiplying U on the right by V^{in} and using the same reasoning as before, we find $U_{ni} = -\bar{U}_{in}$, $1 \leqq i \leqq n-1$. Therefore, if we multiply the last row of U by -1, we obtain a symmetric unitary matrix B for which $B_{nn} < 0$ and such that $\alpha = \sum_{i=1}^{n} B_{ii}$.

In case $U_{nn} = 0$, let k be an integer for which $U_{kn} \neq 0$, and let B be the matrix obtained by multiplying the n-th column of U by $\arg(U_{kn})$ and the n-th row of U by $\arg(U_{nk})$. Then $B_{nk} = B_{kn} > 0$. If $j \neq k$, $1 \leqq j \leqq n-1$, we have

$$B_{jn}B_{kn} = -\sum_{i=1}^{n-1} B_{ji}\bar{B}_{ki} = -\sum_{i=1}^{n-1} \bar{B}_{ij}B_{ik} = \bar{B}_{nj}B_{nk}.$$

Therefore $B_{jn} = \bar{B}_{nj}$. Thus in both cases, we have obtained a symmetric unitary matrix B for which $\alpha = \sum_{i=1}^{n} B_{ii}$ and $B_{nn} \leqq 0$. But the eigenvalues of a symmetric unitary matrix are all ± 1. Also α is equal to the sum of the eigenvalues of B. Since B is not the identity matrix, B_{nn} being non-positive, the trace of B can be at most $n-2$. The proof is complete.

By using Theorem 11, we can prove many curious inequalities in the elements of a unitary matrix. For example, if we apply Theorem 11 to the product

$$\begin{pmatrix} U & 0 \\ 0 & 1 \end{pmatrix} \cdot \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

where U is an arbitrary unitary matrix of order 3, we find

$$|-U_{11}+U_{12}+U_{13}|+|U_{21}-U_{22}+U_{23}|+|U_{31}+U_{32}-U_{33}|\leqq 5.$$

The bound is attained for

$$U=\tfrac{1}{3}\begin{pmatrix}1 & -2 & -2\\ -2 & 1 & -2\\ -2 & -2 & 1\end{pmatrix}.$$

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] G. H. Hardy, J. E. Littlewood, and G. Polya, "*Inequalities*," Cambridge University Press, 1952.

[2] G. Birkhoff, "Three observations on linear algebra," *Universidad Nacional de Tucumán, Revista, Serie* A, vol. 5 (1946), pp. 147-151 (In Spanish).

[3] G. Polya, "Remark on Weyl's note 'Inequalities between the two kinds of eigenvalues of a linear transformation'," *Proceedings of the National Academy of Sciences*, vol. 36 (1950), pp. 49-51.

[4] Ky Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," *ibid.*, vol. 37 (1951), pp. 760-766.

[5] I. Schur, "Über eine klasse von mittelbildungen mit anwendungen auf der determinantentheorie," *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, vol. 22 (1923), pp. 9-20.

[6] A. Horn, "On the eigenvalues of a matrix with prescribed singular values," to appear in *Proceedings of the American Mathematical Society*, vol. 5 (1954).

[7] S. Sherman, "On a theorem of Hardy, Littlewood, Polya, and Blackwell," *Proceedings of the National Academy of Sciences*, vol. 37 (1951), pp. 826-831.

NATURALLY TOTALLY ORDERED COMMUTATIVE SEMIGROUPS.*

By A. H. Clifford.

By a semigroup we mean a set S endowed with an associative binary operation: $a(bc) = (ab)c$. The cancellation law is not assumed. The present paper deals exclusively with commutative semigroups, and the modifier "commutative" will be omitted except in the statements of the theorems.

If a and b are elements of S, we say that a divides b, and write $a \leqq b$, if either $a = b$ or there exists c in S such that $ac = b$. We write $a < b$ if $a \leqq b$ but $b \not\leqq a$. We shall call S *naturally totally ordered* (n. t. o.) if the division relation is a total (i. e. "linear," or "simple") ordering of S: for every pair of elements a, b of S exactly one of the relations $a < b$, $a = b$, $b < a$ holds. If S has an identity element, it is a *linear holoid* as defined by Klein-Barmen [4]. The class of n. t. o. semigroups is, of course, far more restricted than that of linearly ordered semigroups, as defined by O. Nakada [6].

An n. t. o. semigroup S without identity element, and in which the cancellation law holds, is the same thing as a system satisfying O. Hölder's axioms I, III, IV, V, VI ([3], pp. 4-7). Axiom II requires that S be non-discrete in the sense of containing no least element. In [3], pp. 22-27, Hölder shows that if such an S be archimedean, it can be embedded in the additive semigroup P of all positive real numbers. The chief object of the present paper is to establish the analogous result for an archimedean n. t. o. semigroup in which the cancellation law does not hold.

Before describing this result, let us return to the general (non-archimedean) case. In [4], Klein-Barmen determined the structure of all linear holoids S such that either S is finite or else the order type of S is that of the natural numbers. In section 1 below, we give a reduction theorem for the general case, from which his results follow. Suppose we have a set of n. t. o. semigroups S_τ, where the index τ ranges over a totally ordered ("t. o.") set I. The class sum S of all the S_τ becomes a semigroup if we define $a_\sigma b_\tau = b_\tau a_\sigma = b_\tau$ for all $a_\sigma \in S_\sigma$, $b_\tau \in S_\tau$, when $\sigma < \tau$ in I. It is readily seen that S is also n. t. o. We call S the *ordinal sum* of the t. o. set of semigroups S_τ. An n. t. o.

* Received September 16, 1953.

semigroup S which is not the ordinal sum of two or more subsemigroups will be called *ordinally irreducible*. In section 1 we show that every n. t. o. semigroup is uniquely expressible as the ordinal sum of ordinally irreducible n. t. o. semigroups.

This theorem reduces the study of n. t. o. semigroups to that of irreducible ones. The general case appears to be difficult, but one can treat the case in which every element a of S has finite index (i. e. $a^{n+1} = a^n$ for some positive integer n). We shall use the term *segment* to mean an ordinally irreducible, n. t. o., (commutative) semigroup, every element of which has finite index. A segment may also be characterized as an archimedean n. t. o. semigroup in which the cancellation law does not hold (Remark 4 of section 5). In section 2 we show that every segment S has a greatest element, which we call the *ultimate* element of S, and denote by ∞. If we prefer to think of ∞ as the zero element of S, then a segment may be characterized as an n. t. o. semigroup with zero, in which every element is nilpotent. Every discrete segment is a finite cyclic semigroup. Every non-discrete segment S is either dense-in-itself, or is dense-in-itself except for the existence of an immediate predecessor ∞^* of ∞, which we call the *penultimate* element of S.

In section 3 we show that any segment S can be extended in a unique manner to an archimedean n. t. o. semigroup T satisfying the cancellation law. More precisely, T contains an ideal T_∞ such that S is isomorphic with the factor semigroup T/T_∞ in the sense of Rees, [8], p. 389. Using the embedding theorem of Hölder mentioned above, we show in section 4 that S can be embedded in one of the two semigroups $\mathbf{P}_1 = \mathbf{P}/[1]$, $\mathbf{P}^*_1 = \mathbf{P}/(1)$, where $[1]$ and (1) denote the sets of real numbers x satisfying $x \geqq 1$ and $x > 1$, respectively.[1] These two semigroups are, to within isomorphism, the only complete non-discrete segments. $\mathbf{P}_1$ is dense-in-itself, while $\mathbf{P}^*_1$ has the penultimate element $\infty^* = 1$.

The concluding section 5 contains half a dozen remarks, of which we mention one. In Remark 5 it is briefly indicated how a segment S can be completed directly (i. e. without the above extension to T) using Dedekind cuts. Besides being somewhat longer, this treatment suffers from one's having constantly to guard against an interesting but annoying difficulty. If S is a non-discrete segment having a penultimate element ∞^*, the function $z = xy$ fails to be continuous along the curve $z = \infty^*$. It is, however, lower semi-continuous, and the calculus of cuts goes through all right provided we reckon with (closed) lower classes. Thus we must reverse Hölder's definition of a

[1] $\mathbf{P}_1$ is an instance of what Klein-Barmen in [5] calls a "geschlossene Zahlbezirk."

cut, [3], p. 20, and require that the upper class have no least element. The upper classes of S are just the v-ideals of S, as defined in [1], p. 600, and these are therefore not suitable for the completion of S.

Notation. $A \subset B$ means A is a proper subset of B. 0 denotes the empty set. If $A \subseteq B$, $B - A$ denotes the complement of A in B. If $\{A_\tau; \tau \in I\}$ is a set of classes, $\bigcup_{\tau \in I} A_\tau$ denotes their class sum, and $\bigcap_{\tau \in I} A_\tau$ their intersection.

1. Reduction of an n. t. o. semigroup into the ordinal sum of ordinally irreducible n. t. o. semigroups. Let S be an n. t. o. semigroup. A subset $\mathfrak{a}$ of S is an *ideal* of S if $a \in \mathfrak{a}$ and $b \in S$ imply $ab \in \mathfrak{a}$. Thus $\mathfrak{a}$ is an ideal of S if and only if it is an upper class of S in the usual sense ($a \in \mathfrak{a}$ and $a < x$ imply $x \in \mathfrak{a}$). The complement $S - \mathfrak{a}$ of an upper class $\mathfrak{a}$ is a lower class, and the decomposition of S into $\mathfrak{a}$ and $S - \mathfrak{a}$ will be called a *cut* of S. An ideal $\mathfrak{p}$ of S is *prime* if $S - \mathfrak{p}$ is a subsemigroup of S. If a and b are elements of S such that $ab = a$, we shall say that a *absorbs* b. An ideal $\mathfrak{a}$ of S will be called *absorbent* if $a \in \mathfrak{a}$ and $b \notin \mathfrak{a}$ imply $ab = a$.

LEMMA 1.1. *Let $\mathfrak{p}$ be a non-empty, proper, absorbent prime ideal of an n. t. o. semigroup S. Then $\mathfrak{p}$ and $S - \mathfrak{p}$ are non-empty, n. t. o. subsemigroups of S, and S is the ordinal sum of $S - \mathfrak{p}$ and $\mathfrak{p}$ in that order.*

Proof. The subsemigroups $\mathfrak{p}$ and $S - \mathfrak{p}$ are t. o. by the division relation in S. We must show that this ordering is natural in both of them. Let a and b be elements of $\mathfrak{p}$ such that $a < b$ relative to S. Then $ac = b$ for some $c \in S$. Were $c \notin \mathfrak{p}$ we would have $b = ac = a$, contrary to $a < b$. Hence $c \in \mathfrak{p}$, and $a < b$ relative to $\mathfrak{p}$. Similarly, let a and b be elements of $S - \mathfrak{p}$ such that $ac = b$ for some $c \in S$. Were $c \in \mathfrak{p}$ we would have $b = ac = c$, contrary to $b \notin \mathfrak{p}$. Thus $\mathfrak{p}$ and $S - \mathfrak{p}$ are n. t. o. semigroups in their own right. Since they are disjoint, and S is their class sum, and every element of $\mathfrak{p}$ absorbs every element of $S - \mathfrak{p}$, it is clear from the definition that S is their ordinal sum in the order $S - \mathfrak{p}, \mathfrak{p}$.

LEMMA 1.2. *An n. t. o. semigroup is ordinally reducible if and only if it contains a proper, non-empty, absorbent prime ideal.*

Proof. If an n. t. o. semigroup S is the ordinal sum of a set of n. t. o. semigroups S_τ ($\tau \in I$), where I contains more than one element, and if R is any proper, non-empty, upper class of I, then $\mathfrak{p} = \bigcup_{\tau \in R} S_\tau$ is a proper, non-empty, absorbent prime ideal of S. The converse follows from Lemma 1.1.

A prime ideal $\mathfrak{p}$ of an n. t. o. semigroup S will be called *special* if (1) $\mathfrak{p}$

is absorbent, and (2) there exists an absorbent prime ideal $\mathfrak{q}$ such that $\mathfrak{q} \supset \mathfrak{p}$, and such that no absorbent prime ideal $\mathfrak{r}$ exists such that $\mathfrak{q} \supset \mathfrak{r} \supset \mathfrak{p}$. $\mathfrak{q}$ is evidently uniquely determined by $\mathfrak{p}$, and will be called the absorbent prime ideal *associated* with $\mathfrak{p}$. $\mathfrak{q}$ itself need not be special. Condition (2) requires $\mathfrak{p} \neq S$, but we do not exclude the possibility $\mathfrak{p} = 0$. The null set is evidently a special prime ideal of S if and only if S contains a minimal non-empty absorbent prime ideal.

LEMMA 1.3. *Let $\mathfrak{p}$ be a special prime ideal of an n. t. o. semigroup S, and let $\mathfrak{q}$ be its associated absorbent prime ideal. Then $\mathfrak{q} - \mathfrak{p}$ is a non-empty, ordinally irreducible, n. t. o. subsemigroup of S.*

Proof. Applying Lemma 1.1 first to $S - \mathfrak{p}$, and then to the absorbent prime ideal $\mathfrak{q} - \mathfrak{p}$ of the n. t. o. semigroup $S - \mathfrak{p}$, we see that $\mathfrak{q} - \mathfrak{p}$ is a (non-empty) n. t. o. subsemigroup of S. Were it ordinally reducible, it would, by Lemma 1.2, contain a proper, non-empty, absorbent prime ideal P. It is readily seen that $P \cup \mathfrak{p}$ would be an absorbent prime ideal of S between $\mathfrak{p}$ and $\mathfrak{q}$, which is impossible.

THEOREM 1. *Every naturally totally ordered, commutative semigroup is uniquely expressible as the ordinal sum of a totally ordered set of ordinally irreducible such semigroups.*

Proof. Let S be an n. t. o. semigroup. Let I be a set in one-to-one correspondence with the set of special prime ideals of S. Let $\mathfrak{p}_\tau$ be the special prime ideal of S corresponding to the element τ of I. For $\sigma, \tau \in I$, define $\sigma < \tau$ if $\mathfrak{p}_\sigma \supset \mathfrak{p}_\tau$; this is evidently a t. o. of I. Let $\mathfrak{q}_\tau$ be the absorbent prime ideal of S associated with $\mathfrak{p}_\tau$, and let $S_\tau = \mathfrak{q}_\tau - \mathfrak{p}_\tau$. By Lemma 1.3, each S_τ is a non-empty, ordinally irreducible, n. t. o. subsemigroup of S. If $\sigma < \tau$, then $\mathfrak{p}_\sigma \supseteq \mathfrak{q}_\tau \supset \mathfrak{p}_\tau$, so that $S_\tau \subseteq S - \mathfrak{q}_\tau$, while $S_\tau \subseteq \mathfrak{q}_\tau$. Hence $S_\sigma \cap S_\tau = 0$, and, since $\mathfrak{q}_\tau$ is absorbent, every element of S_τ absorbs every element of S_σ. The theorem will follow when we show that every element a of S belongs to some S_τ.

Let $\mathfrak{p}_a$ be the class sum of all the absorbent prime ideals of S not containing a, including the null set. ($\mathfrak{p}_a = 0$ if and only if S contains a minimal non-empty absorbent prime ideal $\mathfrak{q}_0$, and $a \in \mathfrak{q}_0$.) $\mathfrak{p}_a$ is evidently a prime ideal. It is, moreover, absorbent. For suppose $x \in \mathfrak{p}_a$, $y \notin \mathfrak{p}_a$. Then $x \in \mathfrak{p}$ for some absorbent prime ideal $\mathfrak{p}$ not containing a. Since $y \notin \mathfrak{p}$, we conclude $xy = x$.

Let $\mathfrak{q}_a$ be the intersection of all the absorbent prime ideals of S containing a. $\mathfrak{q}_a$ is evidently a prime ideal. To see that it is also absorbent, let $x \in \mathfrak{q}_a$,

$y \notin \mathfrak{q}_a$. Then $y \notin \mathfrak{q}$ for some absorbent prime ideal $\mathfrak{q}$ of S containing a. Since $x \in \mathfrak{q}$, we conclude $xy = x$.

Evidently $\mathfrak{q}_a \supset \mathfrak{p}_a$. There cannot exist an absorbent prime ideal $\mathfrak{r}$ of S between $\mathfrak{p}_a$ and $\mathfrak{q}_a$; for if $a \in \mathfrak{r}$ then $\mathfrak{r} \supseteq \mathfrak{q}_a$, and if $a \notin \mathfrak{r}$ then $\mathfrak{p}_a \supseteq \mathfrak{r}$. Hence $\mathfrak{p}_a$ is a special prime ideal of S, $\mathfrak{q}_a$ is its associated absorbent prime ideal, and $a \in \mathfrak{q}_a - \mathfrak{p}_a$.

To prove the uniqueness, suppose that S is the ordinal sum of a t. o. set of ordinally irreducible n. t. o. semigroups $S'_\alpha (\alpha \in J)$. For each $\alpha \in J$, let $\mathfrak{p}'_\alpha = \bigcup_{\beta > \alpha} S'_\beta$ and $\mathfrak{q}'_\alpha = \mathfrak{p}'_\alpha \cup S'_\alpha$. Then $\mathfrak{p}'_\alpha$ is a special prime ideal of S, and $\mathfrak{q}'_\alpha$ is its associated absorbent prime ideal. For $\mathfrak{p}'_\alpha$ and $\mathfrak{q}'_\alpha$ are evidently absorbent prime ideals of S, and the irreducibility of S'_α prevents the occurrence of an absorbent prime ideal between them. Thus $\mathfrak{p}'_\alpha = \mathfrak{p}_\tau$ for some $\tau \in I$, and $S'_\alpha = S_\tau$. The mapping $\alpha \to \tau$ of J into I must cover I since $\bigcup_{\alpha \in J} S'_\alpha = S$, and is clearly one-to-one and order-preserving.

2. Ordinally irreducible n. t. o. semigroups; segments.

The following lemma is essentially Klein-Barmen's Satz 9 [4], p. 727.

LEMMA 2. 1. *Let e be an idempotent element of an n. t. o. semigroup S. Then the set $\mathfrak{p}$ of all elements x of S satisfying $x > e$ is an absorbent prime ideal of S.*

Proof. $\mathfrak{p}$ is evidently an ideal. If a and b are elements of $S - \mathfrak{p}$, i. e. $a \leqq e$ and $b \leqq e$, then $ab \leqq e^2 = e$, $ab \in S - \mathfrak{p}$, and $\mathfrak{p}$ is prime. Let $a \in \mathfrak{p}$, $b \in S - \mathfrak{p}$, i. e. $a > e$, $b \leqq e$. Then $a = ce$ for some $c \in S$. From $b \leqq e$ we have $e \leqq eb \leqq e^2 = e$, whence $eb = e$. Hence $ab = ceb = ce = a$, and $\mathfrak{p}$ is absorbent.

THEOREM 2. *An ordinally irreducible, naturally totally ordered, commutative semigroup S can contain at most one idempotent element. If it does, then this lone idempotent is the greatest element of S, and absorbs every element of S.*[2]

Proof. Suppose e is an idempotent element of S. By Lemma 1. 2, S cannot contain a non-empty, proper, absorbent prime ideal. The absorbent prime ideal $\mathfrak{p}$ of Lemma 2. 1 is certainly proper, since $e \notin \mathfrak{p}$, and hence it must

[2] Except for the (inconsequential) absence of an identity element in S, Theorem 2 asserts that S is a "fundamental holoid" in the terminology of Klein-Barmen [5]; S is of the "first kind" if it contains the lone idempotent, and otherwise it is of the "second kind."

be empty. Hence e is the greatest element of S. But then, if $a \in S$, $ea \geqq e$, hence $ea = e$, and e absorbs every element a of S.

If an n. t. o. semigroup S contains a greatest element, we shall call it the *ultimate* element of S, and denote it by ∞. Clearly ∞ absorbs every element of S. Conversely, if an element of S absorbs every element of S (i. e. if it is a "zero element" of S) this element exceeds all other elements of S.

Let S be an n. t. o. semigroup, and a any element of S. The subsemigroup C_a of S consisting of the powers of a is called the cyclic subsemigroup of S generated by a. Either $a^n < a^{n+1}$ for all positive integers n, or else there is an n such that $a < a^2 < \cdots < a^n = a^{n+1} = a^{n+2} = \cdots$. In the latter case, as in [1], p. 597, we call n the *index* of a. C_a is itself n. t. o., and if a has index n, a^n is the ultimate element of C_a. C_a is clearly ordinally irreducible.

By the term *segment* we shall mean an ordinally irreducible, n. t. o. semigroup, every element of which has finite index. Let S be a segment. Since S must have at least one idempotent, it has by Theorem 2 exactly one, and this is the ultimate element ∞ of S. If $a \in S$, and n is the index of a, then $a^n = \infty$. If we prefer to regard ∞ as the zero element of S, then every element is nilpotent. Conversely, if S is an n. t. o. semigroup with ultimate element ∞, and if some finite power of each element of S is equal to ∞, then S is a segment.

LEMMA 2.2. *If a and b are elements of a segment S such that $ab = b$, then $b = \infty$.*

Proof. $ab = b$ implies $a^n b = b$ for every n. Taking n to be the index of a, $\infty = \infty \cdot b = b$.

LEMMA 2.3. *If a, b, c are elements of a segment S such that $ab = ac \neq \infty$, then $b = c$.*

Proof. If, say, $b < c$, then $bd = c$ for some $d \in S$. From $ab = ac = abd$ and Lemma 2.2 we would conclude $ab = \infty$.

LEMMA 2.4. *If a and b are elements of a segment S such that $a < b < \infty$, there exists a uniquely determined element b/a of S such that $a(b/a) = b$.*

Proof. Since $a < b$, there exists $c \in S$ such that $ac = b$. If there were another such element d, then from $ac = b = ad$, $b \neq \infty$, and Lemma 2.3, we would conclude $c = d$.

An n. t. o. semigroup S is called *discrete* if it contains a least element.

LEMMA 2.5. *A discrete segment is a finite cyclic semigroup.*

Proof. Let S be a discrete segment, and p its least element. Let a be an arbitrary element $\neq \infty$ of S. Since $p \leqq a$ and $a < p^n = \infty$, where n is the index of p, there is a unique positive integer r such that $p^r \leqq a < p^{r+1}$. If $a \neq p^r$ then $a = p^r b$ for some $b \in S$. Since $b \geqq p$, $a = p^r b \geqq p^{r+1}$, contrary to $a < p^{r+1}$. Hence $a = p^r$, and S is the cyclic semigroup generated by p.

If the ultimate element ∞ of an n. t. o. semigroup S has an immediate predecessor, the latter will be called the *penultimate* element of S, and will be denoted by ∞^*.

LEMMA 2.6. *Let a and b be elements of a non-discrete segment S such that $a < b$. Then either there is an element of S between a and b, or else S has a penultimate element ∞^*, and $a = \infty^*$, $b = \infty$.*

Proof. Assume that no element of S exists between a and b. Since $a < b$, $b = ac$ for some $c \in S$. Since S is non-discrete, there exists $d < c$ in S. Now $a \leqq ad \leqq ac = b$. $a \neq ad$ since otherwise $a = \infty$ by Lemma 2.2, contrary to $a < b$. Hence $ad = ac = b$. Were $b \neq \infty$, this would imply $d = c$ by Lemma 2.3. Hence $b = \infty$, and a is its immediate predecessor.

Lemma 2.6 shows that a non-discrete segment S is dense-in-itself except possibly at ∞. We shall call S *dense* or *dense-except-at-∞* according to whether S does not or does have a penultimate element.

3. Extension of a segment. Let S be a segment. Let Φ be the set of finite sequences $\alpha = \{a_1, \cdots, a_m\}$ of elements $a_i \neq \infty$ of S. We define an evidently associative product in Φ by juxtaposition: if $\alpha = \{a_1, \cdots, a_m\}$ and $\beta = \{b_1, \cdots, b_n\}$, then $\alpha\beta = \{a_1, \cdots, a_m, b_1, \cdots, b_n\}$. The mapping $\alpha \to \Pi\alpha = a_1 a_2 \cdots a_m$ is evidently a homomorphism of Φ onto S: $\Pi(\alpha\beta) = (\Pi\alpha)(\Pi\beta)$. We say that β is a *refinement* of α if there exist elements $\beta_1, \cdots, \beta_m$ of Φ such that $\beta = \beta_1\beta_2 \cdots \beta_m$ and $\Pi\beta_i = a_i$ $(i = 1, \cdots, m)$. The relation "is a refinement of" is transitive and reflexive.

LEMMA 3.1. *If $x_1, x_2, \cdots, x_r$ are elements of a segment S, and if i and j are positive integers $\leqq r$ such that $x_1 x_2 \cdots x_i = x_1 x_2 \cdots x_j \neq \infty$, then $i = j$.*

Proof. Immediate from Lemma 2.2.

LEMMA 3.2. *Let* $\alpha = \{a_1, \cdots, a_m\}$, $\xi = \{x_1, \cdots, x_p\}$, *and* $\eta = \{y_1, \cdots, y_q\}$ *be elements of* Φ. *Then* ξ *and* $\xi\eta$ *cannot both be refinements of* α.

Proof. If ξ is a refinement of α, there exists i such that $a_1 = x_1 x_2 \cdots x_i$. If $\xi\eta$ is also a refinement of α, then either there exists j such that $a_1 = x_1 x_2 \cdots x_j$, or else k such that $a_1 = x_1 x_2 \cdots x_p y_1 y_2 \cdots y_k$. Since $a_1 \neq \infty$, Lemma 3.1 implies that the first case must hold, and $i = j$. Letting $\alpha' = \{a_2, \cdots, a_m\}$ and $\xi' = \{x_{i+1}, \cdots, x_p\}$ we then see that ξ' and $\xi'\eta$ are both refinements of α', and the result follows by induction on the length of α.

LEMMA 3.3. *There exists a single-valued function associating with each pair of elements* α, β *of* Φ *an element* γ *of* Φ *such that exactly one of the following propositions holds:*

(1) γ *is a common refinement of* α *and* β;

(2) γ *is a refinement of* α, *and there exists* $\delta \in \Phi$ *such that* $\gamma\delta$ *is a refinement of* β;

(3) γ *is a refinement of* β, *and there exists* $\delta \in \Phi$ *such that* $\gamma\delta$ *is a refinement of* α.

Remark. γ will be called the *greatest common part* (GCP) *of* α *and* β.

Proof. Let $\alpha = \{a_1, \cdots, a_m\}$, $\beta = \{b_1, \cdots, b_n\}$. Define $c_1 = \min(a_1, b_1)$. Recalling Lemma 2.4, we define elements α_1, β_1, of Φ as follows:

(1) if $a_1 = b_1$, then $\alpha_1 = \{a_2, \cdots, a_m\}$, $\beta_1 = \{b_2, \cdots, b_n\}$;

(2) if $a_1 < b_1$, then $\alpha_1 = \{a_2, \cdots, a_m\}$, $\beta_1 = \{b_1/c_1, b_2, \cdots, b_n\}$;

(3) if $a_1 > b_1$, then $\alpha_1 = \{a_1/c_1, a_2, \cdots, a_m\}$, $\beta_1 = \{b_2, \cdots, b_n\}$.

In each case, $\{c_1\}\alpha_1$ and $\{c_1\}\beta_1$ are refinements of α and β, respectively. We now do exactly the same thing to the pair α_1, β_1: define c_2 to be the lesser of the initial terms of α_1 and β_1, and define $\alpha_2(\beta_2)$ to be the sequence obtained by lopping c_2 off the front end of $\alpha_1(\beta_1)$, as described above. Since $\{c_2\}\alpha_2$ is a refinement α_1, $\{c_1, c_2\}\alpha_2 = \{c_1\}\{c_2\}\alpha_2$ is a refinement of $\{c_1\}\alpha_1$, hence of α; similarly $\{c_1, c_2\}\beta_2$ is a refinement of β. Continuing this process, we construct a sequence of pairs of elements α_i, β_i of Φ, together with a sequence of elements c_i of S, such that (for each $i = 1, 2, \cdots$) $\{c_1, \cdots, c_i\}\alpha_i$ and $\{c_1, \cdots, c_i\}\beta_i$ are refinements of α and β respectively.

To avoid circumlocution, we adjoin to Φ the empty sequence, for the present proof only. The above process must terminate in a finite number p of steps, for at each step at least one of the pair α_i, β_i is shortened. Then α_p,

or β_p, or both, is the empty sequence. Let $\gamma = \{c_1, \cdots, c_p\}$. $\gamma\alpha_p$ and $\gamma\beta_p$ are refinements of α and β, respectively. If both α_p and β_p are empty, we are in Case (1). If just α_p is empty, we are in Case (2) with $\delta = \beta_p$. If just β_p is empty, we are in Case (3) with $\delta = \alpha_p$.

We define $\alpha \sim \beta$ $(\alpha, \beta \in \Phi)$ if α and β have a common refinement ("c. r.") in Φ. Clearly $\sim$ is reflexive and symmetric.

LEMMA 3.4. *The relation $\sim$ is transitive.*

Proof. Let α, β, γ be elements of Φ such that $\alpha \sim \beta$ and $\beta \sim \gamma$. Let ξ be a c. r. of α and β, and let η be a c. r. of β and γ. Let ζ be the GCP (Lemma 3.3) of ξ and η. If Case (1) holds, i. e. ζ is a c. r. of ξ and η, then, since ξ and η are respective refinements of α and γ, we conclude that ζ is a c. r. of α and γ, whence $\alpha \sim \gamma$. We proceed to show that the other cases are impossible. Suppose e. g. that Case (2) held, i. e. ζ is a refinement of ξ, and $\zeta\delta$ is a refinement of η for some $\delta \in \Phi$. Since ξ and η are both refinements of β, we conclude that ζ and $\zeta\delta$ are both refinements of β, contrary to Lemma 3.2. Case (3) is, of course, similar.

Thus $\sim$ is an equivalence relation in Φ. If $\alpha \sim \beta$ then clearly $\alpha\gamma \sim \beta\gamma$ and $\gamma\alpha \sim \gamma\beta$ for every $\gamma \in \Phi$, i. e. $\sim$ is a congruence relation, and the equivalence classes under $\sim$ constitute a semigroup T when multiplied in the usual way. It will be convenient, however, to regard the elements $\alpha, \beta, \cdots$ of Φ as themselves elements of T, with $\sim$ as the equality relation.

LEMMA 3.5. *T is commutative.*

Proof. If $a, b \in S - \{\infty\}$, then $\{a, b\} = \{b, a\}$. This is trivial if $a = b$, and otherwise we may suppose $a > b$. Then $a = bc$ for some $c \in S$. Since $bc = cb = a$, we see that $\{b, c, b\}$ is a c. r. of $\{a, b\}$ and $\{b, a\}$. We may express this as $\{a\}\{b\} = \{b\}\{a\}$, and the lemma follows from the observation that the one-term sequences generate T.

LEMMA 3.6. *The cancellation law holds in T.*

Proof. Let $\alpha = \{a_1, \cdots, a_m\}$, $\beta = \{b_1, \cdots, b_n\}$, $\gamma = \{c_1, \cdots, c_p\}$ be elements of T such that $\alpha\beta \sim \alpha\gamma$. Let $\xi = \{x_1, \cdots, x_r\}$ be a c. r. of $\alpha\beta$ and $\alpha\gamma$. Then we have correspondingly $a_1 = x_1x_2 \cdots x_i$ and $a_1 = x_1x_2 \cdots x_j$ with $1 \leqq i, j \leqq r$. By Lemma 3.1, $i = j$. But then $\xi' = \{x_{i+1}, \cdots, x_r\}$ is a c. r. of $\alpha'\beta$ and $\alpha'\gamma$, where $\alpha' = \{a_2, \cdots, a_m\}$, and we may use induction on the length of α.

LEMMA 3.7. *T does not contain elements α, β such that $\alpha\beta \sim \alpha$.*

Proof. By induction on the length of α, we can show, by an argument similar to that used in the proof of Lemma 3.2, that $\alpha\beta$ and α cannot have a c. r.

COROLLARY. *If α and β are associate elements of T ($\alpha \geqq \beta$ and $\beta \geqq \alpha$) then $\alpha \sim \beta$.*

LEMMA 3.8. *T is naturally totally ordered.*

Proof. Let $\alpha, \beta \in T$, and let γ be their GCP (Lemma 3.3). In Case (1), $\alpha \sim \beta$. In Case (2), $\gamma \sim \alpha$ and $\gamma\delta \sim \beta$, whence $\alpha\delta \sim \beta$, i.e. $\alpha \leqq \beta$ in T. In Case (3), $\beta \leqq \alpha$. The lemma then follows from the corollary to Lemma 3.7.

LEMMA 3.9. (1) *If $\alpha \sim \beta$ then $\Pi\alpha = \Pi\beta$.*

(2) *If $\Pi\alpha = \Pi\beta \neq \infty$, then $\alpha \sim \beta$.*

(3) *If $\alpha \sim \beta\gamma$ then $\Pi\alpha = (\Pi\beta)(\Pi\gamma)$.*

Proof. (1) If ξ is a c. r. of α and β, then $\Pi\alpha = \Pi\xi = \Pi\beta$. (2) Let $c = \Pi\alpha = \Pi\beta$. Then α and β are refinements of $\{c\}$, whence $\alpha \sim \{c\} \sim \beta$. (3) Evidently $\Pi(\beta\gamma) = (\Pi\beta)(\Pi\gamma)$, and $\Pi\alpha = \Pi(\beta\gamma)$ by (1).

An n. t. o. semigroup S is called *archimedean* if, given any two elements $a, b \in S$, there exists a positive integer r such that $a^r \geqq b$.

LEMMA 3.10. *T is archimedean.*

Proof. First let a and b be elements $\neq \infty$ of S. Let r be the index of a, and let $\delta = \{a, a, \cdots, a\}$ with r terms each equal to a. Let $\beta = \{b\}$. Then $\delta \not< \beta$ in T. For otherwise $\gamma\delta \sim \beta$ for some $\gamma \in T$, and $\Pi\beta = (\Pi\gamma)(\Pi\delta)$ by Lemma 3.9 (3). This is impossible, since $\Pi\beta = b \neq \infty$, while $\Pi\delta = a^r = \infty$. Since $\delta = \{a\}^r$, we conclude that $\{a\}^r \geqq \{b\}$.

Now let $\beta = \{b_1, \cdots, b_n\}$. From the foregoing, there exist n positive integers $r_1, \cdots, r_n$ such that $\{a\}^{r_i} \geqq \{b_i\}$, $(i = 1, \cdots, n)$. Letting $r = r_1 + r_2 + \cdots + r_n$, we have $\{a\}^r \geqq \beta$. Finally, if $\alpha = \{a_1, \cdots, a_m\}$ and $\beta = \{b_1, \cdots, b_n\}$, we have $\{a_1\}^r \geqq \beta$ for some r. But $\alpha \geqq \{a_1\}$, whence $\alpha^r \geqq \beta$.

THEOREM 3. *Let S be an ordinally irreducible, naturally totally ordered, commutative semigroup, every element of which has finite index. Then there exists an archimedean, naturally totally ordered, commutative semigroup T, without identity element, and in which the cancellation law holds, such that:*

(A) *T contains an ideal T_∞ such that the factor semigroup T/T_∞, in the sense of Rees, is isomorphic with S; and*

(B) *T is generated by the complement $T - T_\infty$ of T_∞ in T.*

T is uniquely determined to within an isomorphism leaving fixed the elements of $S - \{\infty\}$ if we identify $S - \{\infty\}$ with $T - T_\infty$.

Proof. We take for T the semigroup constructed above. We have already seen that T is an archimedean, n. t. o., commutative semigroup in which the cancellation law holds (Lemmas 3. 5, 3. 6, 3. 8, and 3. 10). The set T_∞ of all $\alpha \varepsilon T$ such that $\Pi\alpha = \infty$ is clearly an ideal of T. Lemma 3. 9 shows that the mapping $\alpha \to \Pi\alpha$, $T_\infty \to \infty$, is an isomorphism between T/T_∞ and S. For we may regard T/T_∞ as composed of the elements of $T - T_\infty$ together with the set T_∞ regarded as a single all-absorbing element. From $\{a_1, \cdots, a_m\} = \{a_1\} \cdots \{a_m\}$ it is clear that $T - T_\infty$ generates T.

To prove the uniqueness, let T' be another semigroup with all the properties stated in the theorem for T, and identify $T' - T'_\infty$ with $S - \{\infty\}$. We shall establish an isomorphism of T onto T' whereby each one-term sequence $\{a\}$ in T maps into the element $a \neq \infty$ of S. We denote the product of two elements α', β' of T' by $\alpha' \circ \beta'$. Then if $a, b \varepsilon S$ we have $a \circ b = ab$ provided $ab \neq \infty$.

Each sequence $\alpha = \{a_1, \cdots, a_m\}$ of Φ determines an element $\alpha' = \Pi'\alpha = a_1 \circ a_2 \circ \cdots \circ a_m$ of T'. If ξ is a refinement of α, we easily see that $\Pi'\xi = \Pi'\alpha$, and consequently $\alpha \sim \beta$ implies $\Pi'\alpha = \Pi'\beta$. Hence $\alpha \to \Pi'\alpha$ is a single-valued mapping of T into T', in fact onto T' by (B), and is evidently a homomorphism. If $\alpha \neq \beta$ in T, then $\Pi'\alpha \neq \Pi'\beta$. For if, say, $\alpha > \beta$, then $\alpha = \beta\gamma$ for some $\gamma \varepsilon T$, and $\Pi'\alpha = (\Pi'\beta)(\Pi'\gamma)$. Since cancellation holds in T', and T' has no identity element, this implies $\Pi'\alpha \neq \Pi'\beta$. Hence $\alpha \to \Pi'\alpha$ is an isomorphism of T onto T', and evidently $\{a\} \to \Pi'\{a\} = a$ for every $a \varepsilon S - \{\infty\}$.

4. Completion of a segment. Let S be an n. t. o. subsemigroup of an n. t. o. semigroup $\mathfrak{S}$. We shall say that S is *everywhere dense* ("e. d.") in $\mathfrak{S}$ if (1) S contains the greatest and the least element of $\mathfrak{S}$, if such exist, and (2) for every pair α, β of distinct elements of $\mathfrak{S}$, either there is an element of S between them, or else there is no element of $\mathfrak{S}$ between them, and in the latter event both α and β belong to S. (The reason for this last condition is given in Remark 6 of section 5.)

Let $\mathfrak{S}$ be a non-discrete segment. Then $\mathfrak{S}$ has no least element, and any

subsemigroup of Σ contains the ultimate element ∞ of Σ. From Lemma 2.6 we conclude:

LEMMA 4.1. *An n. t. o. subsemigroup S of a non-discrete segment Σ is everywhere dense in Σ if and only if there is an element of S between every pair of distinct elements of Σ except the pair ∞^*, ∞, where ∞^* is the penultimate element of Σ (if such exists), and in this event, ∞^* (as well as ∞) belongs to S.*

If Σ is a discrete segment (hence a finite cyclic semigroup), then an n. t. o. subsemigroup of Σ is e. d. in Σ if and only if it coincides with Σ.

An n. t. o. semigroup Σ is called (conditionally) *complete* if every subset of Σ bounded from above has a least upper bound (LUB). This is equivalent to the dual property, that every subset of Σ bounded from below has a greatest lower bound (GLB). Of course, every subset of a segment is bounded from above (by ∞).

In the following theorem, P_1 and P^*_1 denote the semigroups defined in the introduction.

THEOREM 4. *Let S be a non-discrete, ordinally irreducible, naturally totally ordered, commutative semigroup, every element of which has finite index. Then there exists a uniquely determined isomorphism of S onto an everywhere dense subsemigroup of either P_1 or P^*_1 (but not both), the former if S is dense, and the latter if S is dense-except-at-∞.*

Proof. Let T be the semigroup described in Theorem 3. By Hölder's theorem, mentioned in the introduction, there exists an isomorphism ψ of T onto an e. d. subsemigroup ψT of the additive semigroup $\mathbf{P}$ of all positive real numbers. According to Theorem 3, S is isomorphic with the factor semigroup T/T_∞ in the sense of Rees. For convenience, we shall identify $S' = S - \{\infty\}$ with $T - T_\infty$. Let λ be the GLB in $\mathbf{P}$ of ψT_∞. Setting $\phi = \theta\psi$, where θ is the automorphism $\xi \to \lambda^{-1}\xi$ of $\mathbf{P}$ ($\xi \in \mathbf{P}$), ϕ is an isomorphism of T onto an e. d. subsemigroup ϕT of $\mathbf{P}$ such that the GLB in $\mathbf{P}$ of ϕT_∞ is the real number 1. $\phi S'$ is an e. d. subset of the open interval $(0, 1)$, and contains 1 if and only if S' has a greatest element, and the latter must then be the penultimate element ∞^* of S. By Lemma 4.1, ϕS is an e. d. subsemigroup of P_1 or P^*_1, according to whether S does not or does have a penultimate element. (Of course we extend ϕ so that it maps the ultimate element of S onto that of P_1 or P^*_1.)

Since an e. d. subsemigroup of a segment Σ has a penultimate element if and only if Σ has one, it is clear that S cannot be isomorphic with e. d.

subsemigroups of both P_1 and P^*_1. Suppose ϕ_1 and ϕ_2 are two isomorphisms of S onto e. d. subsemigroups $S_1 = \phi_1 S$, $S_2 = \phi_2 S$ of P_1 (or P^*_1). Then $\phi_2\phi_1^{-1}$ is an isomorphism of S_1 onto S_2. As a continuous function defined on an e. d. subset of the interval $(0, 1)$, $\phi_2\phi_1^{-1}$ can be extended in the usual way to a continuous function η on P_1 (or P^*_1) to itself, and η is evidently an automorphism. But the only such is the identity mapping. This can be seen, for example, by noting that each element of the e. d. subsemigroup of P_1 (or P^*_1) generated by the numbers 2^{-i} $(i = 1, 2, \cdots)$ must be left fixed. Hence $\phi_1 = \phi_2$.

COROLARY. P_1 *and* P^*_1 *are, to within isomorphism, the only complete non-discrete segments.*

Proof. If S is a complete, non-discrete segment, and ϕ is an isomorphism of S onto an e. d. subsemigroup ϕS of P_1 or P^*_1, then ϕS is also complete, and so must coincide with P_1 or P^*_1 (as the case may be).

5. Concluding remarks.

Remark 1. Theorem 1 still holds if we delete the word "naturally." The proof becomes almost trivial when freed of the necessity of proving that such and such a t. o. is natural. The notions of "upper class" and "ideal" are no longer synonymous, and we must use the former. The result is of limited value because the corresponding condition of ordinal irreducibility is far weaker than before. It is satisfied by Examples 3, 5, 8, 10, 12, and 14 in [7].

Remark 2. We may obtain a partial reduction of an n. t. o. semigroup S by means of the set E of idempotent elements of S. E is an n. t. o. subsemigroup of S with $ef = \max(e, f)$ for all $e, f \in E$. Let J be the set of all cuts of the t. o. set E, as defined in section 1. For each $\alpha \in J$, let L_α and R_α be the lower and upper classes, respectively, of α. Define S_α to be the set of all $x \in S$ satisfying (i) $e \in L_\alpha$ implies $e < x$, and (ii) $f \in R_\alpha$ implies $x \leqq f$. Then it is readily seen from Lemma 2. 1 that each S_α is a (possibly empty) n. t. o. subsemigroup of S, and that S is their ordinal sum. In fact, Lemma 2. 1 states that the set $\mathfrak{p}_e$ of all $x \in S$ satisfying $x > e$ is an absorbent prime ideal. So also are $\mathfrak{p}_\alpha = \bigcup_{f \in R_\alpha} \mathfrak{p}_f$ and $\mathfrak{q}_\alpha = \bigcap_{e \in L_\alpha} \mathfrak{p}_e$; and $S_\alpha = \mathfrak{q}_\alpha - \mathfrak{p}_\alpha$. Each S_α contains at most one idempotent element e_α, and if so, e_α is the ultimate element of S_α (cf. Theorem 2). If every element of S has finite index, then the S_α are ordinally irreducible.

Remark 3. Let S be an n. t. o. semigroup in which the cancellation law holds. If $ab = b$ for some $a, b \in S$, we see at once that a is the identity element of S. Hence S is ordinally irreducible unless it has an identity element i, and in the latter event it is the ordinal sum of the irreducible n. t. o. semigroups $\{i\}$ and $S - \{i\}$. If we order the quotient group G of S in the usual way (cf. [6], Theorem 5, p. 182), G becomes a t. o. commutative group, and S consists of all elements $\geqq i$ (or $> i$), where i is the identity element of G. Conversely, the set of all non-negative (or of all positive) elements of a t. o. abelian group is an n. t. o. semigroup with cancellation.

Remark 4. We show here that *an archimedean n. t. o. semigroup S, in which the cancellation law does not hold, is a segment.* For then a, b, c exist in S such that $ab = ac$ and $b < c$. $bd = c$ for some $d \in S$, and $ab = abd$. Writing $f = ab$, we have $f = df = d^2f = \cdots$. Since $d^n \geqq f$ for some n, we have $f = d^n f \geqq f^2$, whence $f^2 = f$. For any $a \in S$ we have $f^n \geqq a$ for some n, hence $f \geqq a$, i. e. f is the ultimate element ∞ of S. For any $a \in S$ we also have $a^n \geqq \infty$ for some n, i. e. $a^n = \infty$, and we have observed in section 2 that such a semigroup is a segment.

If the cancellation law holds, then a complete n. t. o. semigroup is archimedean. This is not so if the cancellation law does not hold. For example, let S be the ordinal sum of a conditionally complete t. o. set of complete segments S_τ $(\tau \in I)$, such that S_τ is discrete if τ has no immediate predecessor in I. In fact this is the most general complete n. t. o. semigroup if we allow the cancellation law to hold in S_μ, where μ is the greatest element in I (if such exists).

Remark 5. The completion of a segment S can be carried out by means of cuts in the t. o. set S. The *closure* $\mathfrak{a}'$ of a lower class $\mathfrak{a}$ of S is defined as follows: if $\mathfrak{a}$ has a LUB a_0 in S, then $\mathfrak{a}' = \mathfrak{a} \cup \{a_0\}$, and otherwise $\mathfrak{a}' = \mathfrak{a}$. $\mathfrak{a}$ is *closed* if $\mathfrak{a}' = \mathfrak{a}$. Let $\mathfrak{S}$ be the set of closed lower classes of S. Define $\mathfrak{a} \circ \mathfrak{b} = (\mathfrak{a}\mathfrak{b})'$, where $\mathfrak{a}, \mathfrak{b} \in \mathfrak{S}$, and $\mathfrak{a}\mathfrak{b}$ denotes the set of all ab with $a \in \mathfrak{a}$, $b \in \mathfrak{b}$. $\mathfrak{S}$ becomes thereby a complete n. t. o. semigroup. For $a \in S$, the set (a) of all $x \leqq a$ is called the *principal* closed lower class defined by a. The set $\mathfrak{S}_0$ of all such is an e. d. subsemigroup of $\mathfrak{S}$ isomorphic with S. $\mathfrak{S}$ is unique in the following sense: if $\mathfrak{T}$ is any complete segment containing S as an e. d. subsegment, there exists an isomorphism between $\mathfrak{T}$ and $\mathfrak{S}$ whereby $a \leftrightarrow (a)$ for all $a \in S$. The corollary to Theorem 4 is shown by constructing an e. d. subsemigroup Q of the given complete segment S which is isomorphic with

the e. d. subsemigroup Q' of P_1 or P^*_1 (depending on the type of S) generated by the numbers 2^{-i} $(i = 1, 2, \cdots)$, and then using the above uniqueness theorem.

All this sounds very simple and quite the usual thing, but the dual procedure breaks down. The binary operation $\circ$ defined dually for closed upper classes is no longer associative. For example, take the subsemigroup Q' of P^*_1 mentioned above. Let (a) denote the set of all $x \in Q'$ satisfying $x \geqq a$ (even if $a \notin Q'$). Take $\mathfrak{a} = (1/6)$, $\mathfrak{b} = (1/3)$, $\mathfrak{c} = (1/2)$. Then

$$\mathfrak{a} \circ \mathfrak{b} = (1/2), \qquad (\mathfrak{a} \circ \mathfrak{b}) \circ \mathfrak{c} = \{\infty^*, \infty\};$$

$$\mathfrak{b} \circ \mathfrak{c} = (5/6), \qquad \mathfrak{a} \circ (\mathfrak{b} \circ \mathfrak{c}) = \{\infty\}.$$

Proof of the associativity of $\circ$ depends upon the following lemma: a segment S is *lower semi-continuous* in the sense that if A and B are subsets of S with LUB's a_0 and b_0, respectively, then $a_0 b_0$ is the LUB of AB. The dual of this lemma breaks down when S has a penultimate element ∞^*, and $a_0 b_0 = \infty^*$: let A be the set of all $a > a_0$, and B that of all $b > b_0$; then $ab = \infty$ for every $a \in A$, $b \in B$, so that the LUB of AB is ∞ $(\neq a_0 b_0)$.

On the other hand, multiplication of v-ideals *is* associative ([1], p. 601). The difficulty here is that every upper class of a segment S is a v-ideal, and consequently the semigroup of all v-ideals is not n. t. o. For if $\mathfrak{a}' = (a)$, and $\mathfrak{a} = \mathfrak{a}' - \{a\}$, there cannot exist a v-ideal $\mathfrak{b}$ such that $\mathfrak{a}' = \mathfrak{a}\mathfrak{b}$.

Remark 6. Let S be an n. t. o. subsemigroup of a complete n. t. o. semigroup $\mathfrak{S}$. We shall say that $\mathfrak{S}$ is a *normal completion* of S if every element of $\mathfrak{S}$ is the GLB of a subset of S. This is concordant with the definition of "normal ideal arithmetic" given in [1], p. 605. Condition (N_1), that a divides b relative to S $(a, b \in S)$ if and only if a divides b relative to $\mathfrak{S}$, is satisfied because S is n. t. o. Condition (N_2) requires that every element of $\mathfrak{S}$ be the GCD (= GLB) of a *finite* subset of S. But since the divisor chain condition was assumed in [1] to hold in $\mathfrak{S}$, the qualification "finite" could have been omitted.

If $\mathfrak{S}$ is a normal completion of a segment S, then $\mathfrak{S}$ is easily seen to be a segment also. S is discrete if and only if $\mathfrak{S}$ is discrete, and then $S = \mathfrak{S}$; otherwise S and $\mathfrak{S}$ are either both dense, or both dense-except-at-∞. If S is a subsemigroup of a complete segment $\mathfrak{S}$, then $\mathfrak{S}$ is a normal completion of S if and only if S is e. d. in $\mathfrak{S}$. Theorem 4 implies that every segment possesses a unique normal completion.

If $\mathfrak{S}$ is a normal completion of a segment S, then every element of $\mathfrak{S}$ is also the LUB of a subset of S. The following example shows that the converse

is not true. Let $\Sigma = P^*_1$, and let S be the subsemigroup of P^*_1 consisting of all rational multiples of $2^{\frac{1}{2}}$, together with ∞. Every element of P^*_1 is the LUB of a subset of S, but 1 $(= \infty^*)$ is not the GLB of a subset of S. S would be e. d. in P^*_1 were it not that $1 \notin S$. This example explains why we required $\infty^* \varepsilon S$ in the definition of "everywhere dense," in section 4. If we regarded S as e. d. in P^*_1, then (since S itself is dense) the uniqueness clause in Theorem 4 would not hold.

Remark 7 (added January 7, 1954). In a recent conversation, Professor A. M. Gleason told me that the results of the present paper are closely related to certain work of his own, namely the developments of section 2 of reference [2], especially Lemma 3. In fact, although he did not publish it, Professor Gleason was aware of both Theorem 1 and the Corollary to Theorem 4 in the case S is compact and connected in its order topology.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. H. Clifford, "Arithmetic and ideal theory of commutative semigroups," *Annals of Mathematics*, vol. 39 (1938), pp. 594-610.

[2] A. M. Gleason, "Arcs in locally compact groups," *Proceedings of the National Academy of Sciences*, vol. 36 (1950), pp. 663-667.

[3] O. Hölder, "Die Axiome der Quantität und die Lehre vom Mass," *Berichte über die Verhandlungen der Königlich Sachsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Classe*, vol. 53 (1901), pp. 1-64.

[4] F. Klein-Barmen, "Über gewisse Halbverbande und kommutative Semigruppen II," *Mathematische Zeitschrift*, vol. 48 (1942-3), pp. 715-734.

[5] ———, "Ein Beitrag zur Theorie der linearen Holoide," *Mathematische Zeitschrift*, vol. 51 (1947-9), pp. 355-366.

[6] O. Nakada, "Partially ordered abelian semigroups. I. On the extension of the strong partial order defined on abelian semigroups," *Journal of the Faculty of Science*, Hokkaido University, Sapporo, Japan, vol. 11 (1951), pp. 181-189.

[7] ———, "Partially ordered abelian semigroups. II. On the strongness of the linear order defined on abelian semigroups," *ibid.*, vol. 12 (1952), pp. 73-86.

[8] D. Rees, "On semi-groups," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 387-400.

SIMPLE BASIC SETS OF POLYNOMIALS.*

By M. N. MIKHAIL.

---

1. Introduction. The subject of basic sets of polynomials was introduced by J. M. Whittaker. For such a set $\{p_n(z)\}$ any polynomial and in particular the polynomial z^n admits a unique finite representation of the form:

$$z^n = \Sigma_i \pi_{ni} p_i(z) = \Sigma_i \pi_{ni} \Sigma_j p_{ij} z^j,$$

i. e. $\{\pi_{ij}\}$, $\{p_{ij}\}$ are the basic sets of operators and coefficients, respectively, associated with the set $\{p_n(z)\}$.

Thus any set of polynomials $\{p_n(z)\}$, where $p_n(z)$ is of degree n, is basic and is said to be *simple*.

Associated with any function $f(z)$ regular about the origin, there is a basic series $\Sigma_n \Pi_n f(0) p_n(z)$, where the operator Π_n is given by:

$$\Pi_n = \pi_{on} + \pi_{1n} d/dz + \pi_{2n} d^2/dz^2 + \cdots.$$

The series is said to be *effective* on $|z| = R$, if it represents in and on $|z| = R$ every function which is regular in and on $|z| = R$.

The reciprocal set of polynomials of $\{p_n(z)\}$ is denoted by $\{p_n(z)\}^*$ and is defined by: $z^n = \Sigma_i p_{ni} \Sigma_j \pi_{ij} z^j$.

The product set $\{u_n(z)\} = \{p_n(z)\}\{q_n(z)\}$ is defined by the relations: $z^n = \Sigma_i \mu_{ni} u_i(z) = \Sigma_i \mu_{ni} \Sigma_j u_{ij} z^j$, where $u_{ij} = \Sigma_h p_{ih} q_{hj}$, $\mu_{ij} = \Sigma_h \lambda_{ih} \pi_{hj}$.

The quotient set $\{u_n(z)\} = \{p_n(z)\}\{q_n(z)\}^*$ is defined accordingly by: $u_{ij} = \Sigma_h p_{ih} \lambda_{hj}$, $\mu_{ij} = \Sigma_h q_{ih} \pi_{hj}$.

According to Theorem 1 [1], the order ω of the basic set of polynomials $\{p_n(z)\}$ is defined as follows: $\omega = \limsup_n (n \log n)^{-1} \log \max_{ij} |\pi_{ni}| |p_{ij}|$.

2. In this paper, I investgate the order and the effectiveness of the reciprocal, product, and quotient sets of some class of basic sets of polynomials. My treatment deals with non-normalised simple sets, so that my results imply the results in [2, 3, 4, 5 and 6] for normalised simple sets. This investigation has been made possible by applying some of my results [1].

* Received April 15, 1953.

3. I prove in this paper the following theorems:

THEOREM 1. *Let* $\{p_n(z)\}$ *be a simple basic set of polynomials of order* ω *and such that* $|p_{nn}| = O(n^h)$ $(h, finite)$; *then its reciprocal set is of order* ω^* *such that* $\frac{1}{2}\omega \leqq \omega^* \leqq 2\omega$.

THEOREM 2. *Let* $\{p_n(z)\}$, $\{q_n(z)\}$ *be two simple basic sets of polynomials such that* $|q_{nn}| = O(n^h)$ $(h, finite)$, *and of orders* ω_p, ω_q *respectively and* $\omega_p > 2\omega_q \geqq 0$. *Then the product set* $\{u_n(z)\} = \{p_n(z)\}\{q_n(z)\}$ *is of order* ω_u *such that*: $\omega_p + 2\omega_q \geqq \omega_u \geqq \omega_p - 2\omega_q$.

THEOREM 3. *Let* $\{p_n(z)\}$, $\{q_n(z)\}$ *be two simple basic sets of polynomials such that* $|p_{nn}| = |q_{nn}| = O(n^h)$ $(h, finite)$ *and of orders* ω_p, ω_q *respectively, and* $\omega_q > 2\omega_p \geqq 0$. *Then the order of the product set* $\{u_n(z)\} = \{p_n(z)\}\{q_n(z)\}$ *is* ω_u *such that*: $\omega_p + 2\omega_q \geqq \omega_u \geqq \frac{1}{2}\omega_q - \omega_p$.

THEOREM 4. *Let* $\{p_n(z)\}$ *be a simple basic set of polynomials such that* $|p_{nn}| = O(n^h)$ $(h, finite)$ *and* $C \leqq 1$. *Then both the set and its reciprocal set are effective in the same region.*

THEOREM 5. *Let* $\{p_n(z)\}$, $\{q_n(z)\}$ *be two simple basic sets of polynomials such that* $|q_{nn}| = O(n^h)$ $(h, finite)$. *Then both the product set* $\{p_n(z)\}\{q_n(z)\}$ *and the quotient set* $\{p_n(z)\}\{q_n(z)\}^*$ *are effective on* $|z| = R$, *where* $R \geqq \max(b_p; b_q)$.

Proof of Theorem 1. We have

$$\begin{aligned}\omega^* &= \limsup_n (n\log n)^{-1}\log \max_{ij} |p_{ni}|\,|\pi_{ij}|, \quad \text{by Theorem 1 of [1]}\\ &= \limsup_n (n\log n)^{-1}\log |p_{ni(n)}|\,|\pi_{i(n)j(n)}| \quad (n \geqq i(n) \geqq j(n)).\\ &= \limsup_n \{(n\log n)^{-1}\log |p_{ni(n)}| + (n\log n)^{-1}\log |\pi_{i(n)j(n)}|\}.\end{aligned}$$

Since $|p_{nn}| = O(n^h)$, we have:

$$\lim (n\log n)^{-1}\log |\pi_{nn}| = 0 \text{ and } \lim (n\log n)^{-1}\log |p_{jj}| = 0 \quad (n\to\infty).$$

Therefore, we can write

$$\begin{aligned}\omega^* = \limsup_n (n\log n)^{-1}\log |\pi_{nn}|\,|p_{ni(n)})| \\ + \limsup_n (n\log n)^{-1}\log |\pi_{i(n)j(n)}|\,|p_{j(n)j(n)}|.\end{aligned}$$

Then $\omega^* \leqq \omega + \omega = 2\omega$. In a similar way, we get $\omega \leqq \omega^* + \omega^* = 2\omega^*$. So $\frac{1}{2}\omega \leqq \omega^* \leqq 2\omega^*$.

Proof of Theorem 2. (a) Since both sets are simple, $a_p = a_q = 1$. By Theorem 3 of [1], we get

$$\omega_{pq} \leqq \limsup_n (n \log n)^{-1} \log \max_m |\lambda_{nm}| + \omega_p + |\limsup_n (n \log n)^{-1} \log \max_j |q_{nj}|,$$

but

$$\limsup_n (n \log n)^{-1} \log \max_m |\lambda_{nm}| = \limsup_n (n \log n)^{-1} \log |\lambda_{nm(n)}| = \limsup_n (n \log n)^{-1} \log |\lambda_{nm(n)}| |q_{m(n)m(n)}|,$$

because $|q_{nn}| = O(n^h)$ for all n. So, by Theorem 1 of [1],

$$\limsup_n (n \log n)^{-1} \log \max_m |\lambda_{nm}| \leqq \omega_q.$$

Also

$$|\limsup_n (n \log n)^{-1} \log \max_j |q_{nj}|| = |\limsup_n (n \log n)^{-1} \log |q_{nj(n)}| = |\limsup_n (n \log n)^{-1} \log |\lambda_{nn}| |q_{nj(n)}||$$

because, since the set is simple, $|\lambda_{nn}| = |q_{nn}|$. Then

$$|\limsup_n (n \log n)^{-1} \log \max_j |q_{nj}|| \leqq \omega_q,$$

by Theorem 1 of [1]. Hence $\omega_{pq} \leqq \omega_p + 2\omega_q$.

(b) Since

$\{u_n(z)\} = \{p_n(z)\}\{q_n(z)\}$, it is seen that $\{p_n(z)\} = \{u_n(z)\}\{q_n(z)\}^*$.

Since $a_p = a_q = 1$, by Theorem 4 of [1], we get

$$\omega_p \leqq \limsup_n (n \log n)^{-1} \log \max_m |q_{nm}| + \omega_u + |\limsup_n (n \log n)^{-1} \log \max_j |\lambda_{nj}||.$$

But

$$\limsup_n (n \log n)^{-1} \log \max_m |q_{nm}| = \limsup_n (n \log n)^{-1} \log |q_{nm(n)}| = \limsup_n (n \log n)^{-1} \log |\lambda_{nn}| |q_{nm(n)}| \leqq \omega_q,$$

by Theorem 1 of [1]. And

$$\limsup_n (n \log n)^{-1} \log \max_j |\lambda_{nj}| = \limsup_n (n \log n)^{-1} \log |\lambda_{nj(n)}| = \limsup_n (n \log n)^{-1} \log |\lambda_{nj(n)}| |q_{j(n)j(n)}| \leqq \omega_q,$$

by Theorem 1 of [1]. Hence

$$\omega_p \leqq \omega_q + \omega_u + \omega_q = 2\omega_q + \omega_u,$$

i. e., $\omega_u \geqq \omega_p - 2\omega_q$. Thus the theorem is proved.

Proof of Theorem 3. (a) It has been proved in the last theorem that $\omega_u \leqq \omega_p + 2\omega_q$.

(b) Since $\{u_n(z)\} = \{p_n(z)\}\{q_n(z)\}$, then $\{q_n(z)\} = \{p_n(z)\}^*\{u_n(z)\}$. Hence from (a), we obtain $\omega_q \leqq \omega_p{}^* + 2\omega_u$. But from Theorem 1, we know that $\omega_p{}^* \leqq 2\omega_p$, so that $\omega_q \leqq 2\omega_p + 2\omega_u$, i. e., $\omega_u \geqq \frac{1}{2}\omega_q - \omega_p$. Thus

$$\omega_p + 2\omega_q \geqq \omega_u \geqq \tfrac{1}{2}\omega_q - \omega_p.$$

Proof of Theorem 4. By Theorem 7 of [1], $\{p_n(z)\}$ is effective on $|z| = R$, where $R \geqq b$, and $\{p_n(z)\}^*$ is effective on $|z| = R$ where $R \geqq b^*$ if $C^* \leqq 1$. Thus, the proof of Theorem 4 will be given in two parts: (a) $C^* \leqq 1$, (b) $b^* = b$.

(a) $$C^* = \sup_{d(n)}[\limsup_n\{\sup_i |p_{ni}||\pi_{id(n)}|\}^{1/n}],$$
$$= \sup_{d(n)}[\limsup_n\{|p_{ni(n)}||\pi_{i(n)d(n)}|\}^{1/n}].$$
$$(n \geqq i \geqq d(n)).$$

Since $|p_{nn}| = O(n^h)$ and the set is simple, then

$$C^* = \sup_{d(n)}[\limsup_n\{\{|\pi_{nn}||p_{ni(n)}|\}^{1/n} \times \{\{|\pi_{i(n)d(n)}||p_{d(n)d(n)}|\}^{1/i(n)}\}^{i(n)/n}\}].$$

But, as $n \to \infty$, $\lim i(n)/n = \lim d(n)/n = 1$. Hence $C^* \leqq C \cdot C \leqq 1$ (because $C \leqq 1$).

(b) $$b^* = \sup_{s(n)}[\limsup_n\{\sup_i |p_{ni}||\pi_{is(n)}|\}^{1/(n-s(n))}$$
$$= \sup_{s(n)}[\limsup_n\{|p_{ni(n)}||\pi_{i(n)s(n)}|\}^{1/(n-s(n))}].$$

Since, as $n \to \infty$, $\lim i(n)/n \leqq 1$, there are two sub-cases:

(1) $\lim i(n)/n = 1$, (2) $\lim i(n)/n < 1$.

(1) If $\lim i(n)/n = 1$, we have

$$b^* = \sup_{s(n)}[\limsup_n\{\{|\pi_{nn}||p_{ni(n)}|\}^{1/n}\}^{n/(n-s(n))} \times \limsup_n\{\{|\pi_{i(n)s(n)}||p_{s(n)s(n)}|\}^{1/(i(n)-s(n))}\}^{(i(n)-s(n))/(n-s(n))}],$$

so that $b^* \leqq 1 \times b$ (because $C \leqq 1$).

(2) If $\lim i(n)/n < 1$, we get

$$b^* = \sup_{s(n)}[\limsup_n\{\{|\pi_{nn}||p_{ni(n)}|\}^{1/(n-i(n))}\}^{(n-i(n))/(n-s(n))} \times \limsup_n\{\{|\pi_{i(n)s(n)}||p_{s(n)s(n)}|\}^{1/(i(n)-s(n))}\}^{(i(n)-s(n)/(n-s(n))}$$

so that b^* does not exceed

$$\sup_{s(n)}[\limsup_n\{b(s(n))\}^{(n-i(n))/(n-s(n))}\cdot\{b(s(n))\}^{(i(n)-s(n))/(n-s(n))}.$$

Then $b^* \leqq b$. Similarly it can be proved that $b \leqq b^*$. Hence $b^* = b$, and the result of the theorem follows.

Proof of Theorem 5. The proof of this theorem is in two parts: (a) $C_{pq} \leqq 1$, (b) $b_{pq} = \max(b_p; b_q)$.

(a) $$C_{pq} = \sup_{d(n)}[\limsup_n\{\sup_{mix}|\lambda_{nm}||\pi_{mi}||p_{ix}||q_{xd(n)}|\}^{1/n}]$$
$$= \sup_{d(n)}[\limsup_n\{|\lambda_{nm(n)}||\pi_{m(n)i(n)}||p_{i(n)x(n)}||q_{x(n)d(n)}|\}^{1/n}].$$

Since $|q_{nn}| = O(n^h)$ and the set is simple, then

$$C_{pq} = \sup_{d(n)}[\limsup_n\{\{|\lambda_{nm(n)}||q_{m(n)m(n)}|\}^{1/n}$$
$$\times\{\{|\pi_{m(n)i(n)}||p_{i(n)x(n)}|\}^{1/m(n)}\}^{m(n)/n}\{\{|\lambda_{x(n)x(n)}||q_{x(n)d(n)}|\}^{1/x(n)}\}^{x(n)/n}\}]$$

since $n \geqq m \geqq i \geqq x \geqq d(n)$ and $C_p \leqq 1$, $C_q \leqq 1$. Thus $C_{pq} \leqq 1$.

(b) $$b_{pq} = \sup_{s(n)}[\limsup_n\{|\lambda_{nm(n)}||\pi_{m(n)i(n)}||p_{i(n)x(n)}||q_{x(n)s(n)}|\}^{1/(n-s(n))}.$$

Then

(5.1) $$b_{pq} = \sup_{s(n)}[\limsup_n\{|\lambda_{nm(n)}||p_{m(n)m(n)}|\}^{1/(n-s(n))}$$
$$\times\limsup_n\{|\pi_{m(n)i(n)}||p_{i(n)s(n)}|\}^{1/(n-s(n))}$$
$$\times\limsup_n\{|\lambda_{x(n)x(n)}||q_{x(n)s(n)}|\}^{1/(n-s(n))},$$

since $m(n) \geqq i(n) \geqq x(n) \geqq s(n)$ and, as $n \to \infty$,

$$1 \geqq \lim m(n)/n \geqq \lim i(n)/n \geqq \lim x(n)/n.$$

There are many cases for the value of b_{pq} due to the possible values of $m(n)$, $i(n)$, $x(n)$ and $s(n)$. I shall be satisfied to treat here two cases because the other cases can be treated in a similar way.

(1) $\lim m(n)/n = 1$, $\quad 1 > \lim x(m)/n > \lim s(n)/n$,

then (5.1) gives

$$b_{pq} = \sup_{s(n)}[\limsup_n\{\{|\lambda_{nm(n)}||q_{m(n)m(n)}|\}^{1/n}\}^{n/(n-s(n))}$$
$$\times\limsup_n\{\{|\pi_{m(n)i(n)}||p_{i(n)x(n)}|\}^{1/(m(n)-x(n))}\}^{(m(n)-x(n))/(n-s(n))}$$
$$\times\limsup_n\{\{|\lambda_{x(n)x(n)}||q_{x(n)s(n)}|\}^{1/(x(n)-s(n))}\}^{(x(n)-s(n)/(n-s(n))}].$$
$$\leqq \sup_{s(n)}[\limsup_n\{1\times\{b_p(s(n))\}^{(m(n)-x(n))/(n-s(n))}$$
$$\times\{b_q(s(n))\}^{(x(n)-s(n))/(n-s(n))}\}]$$

(because $C_q \leqq 1$). Hence $b_{pq} \leqq \max(b_p ; b_q)$.

(2) $1 > \lim m(n)/n = \lim x(n)/n > \lim s(n)/n$, then (5. 1) gives

$$b_{pq} = \sup_{s(n)} [\lim \sup_n \{ \{ \{ | \lambda_{nm(n)} | \, | \, q_{m(n)m(n)} | \}^{1/(n-m(n))} \}^{(n-m(n))/(n-s(n))}$$
$$\times \{ \{ | \pi_{m(n)i(n)} | \, | \, p_{i(n)x(n)} | \}^{1/m(n)} \}^{m(n)/(n-s(n))}$$
$$\times \{ \{ | \lambda_{x(n)x(n)} | \, | \, q_{x(n)s(n)} | \}^{1/(x(n)-s(n))} \}^{(x(n)-s(n))/(n-s(n))} \}]$$
$$\leqq \sup_{s(n)} [\lim \sup_n \{ \{ b_q(s(n)) \}^{(n-m(n))/(n-s(n))}$$
$$\times 1 \times \{ b_q(s(n)) \}^{(x(n)-s(n))/(n-s(n))} \}]$$

(because $C_p \leqq 1$). Hence $b_{pq} \leqq \max (b_p ; b_q)$. In a similar way, we find that in every case

$$b_{pq} \leqq \max (b_p ; b_q). \tag{5. 2}$$

Since always (i. e. in every possible case)

$$b_{pq} \geqq \sup_{s(n)} [\lim \sup_n \{ | \lambda_{nm} | \, | \, \pi_{mi} | \, | \, p_{ix} | \, | \, q_{xs(n} | \}^{1/(n-s(n))}],$$

for all values of m, i and x, consider the two cases (i) $b_p \geqq b_q$ or (ii) $b_q > b_p$.

(i) If $b_p \geqq b_q$, then b_{pq} is not less than

$$\sup_{s(n)} [\lim \sup_n \{ \sup_i \{ | \lambda_{nn} | \, | \, \pi_{ni} | \, | \, p_{is(n)} | \, | \, q_{s(n)s(n)} | \}^{1/(n-s(n))} \}].$$

Then $b_{pq} \geqq b_p$.

(ii) If $b_q > b_p$, then b_{pq} is not less than

$$\sup_{s(n)} [\lim \sup_n \{ \sup_i \{ | \lambda_{ni} | \, | \, \pi_{ii} | \, | \, p_{is(n)} | \, | \, q_{s(n)s(n)} | \}^{1/(n-s(n))} \}].$$

Then $b_{pq} \geqq b_q$.

Combining the results of (i) and (ii), we get

$$b_{pq} \geqq \max (b_p ; b_q). \tag{5. 3}$$

Hence in view of (5. 2) and (5. 3), we have $b_{pq} = \max(b_p ; b_q)$. Then by Theorem 9 of [1], the product set $\{p_n(z)\}\{q_n(z)\}$ is effective on $| z | = R$, where $R \geqq \max (b_p ; b_q)$.

In a similar way, it can be proved that the quotient set $\{p_n(z)\}\{q_n(z)\}^*$ is effective on $| z | = R$, where $R \geqq \max (b_p ; b_q)$.

INSTITUTE OF EDUCATION FOR GIRLS,
ZAMALEK, CAIRO, EGYPT.

REFERENCES.

[1] M. N. Mikhail, "On the mode of increase and the effectiveness of basic sets of polynomials, their product set and their quotient sets," *Duke Mathematical Journal*, vol. 20 (1953), pp. 459-480.

[2] M. Nassif, "On the order of product sets of polynomials," *Proceedings of the Mathematical and Physical Society of Egypt*, vol. 3 (1946), pp. 43-47.

[3] ———, "On the product of simple series of polynomials," *Journal of London Mathematical Society*, vol. 22 (1947), pp. 257-261.

[4] M. T. Eweida, *Thesis*, Liverpool University (1948).

[5] M. Mursi and Makar, "Sur la base inverse d'une base de polynomes," *Bulletin des Sciences Mathématiques*, ser. (2), vol. 71 (1947), pp. 47-51.

[6] J. M. Whittaker, *Sur les séries de base de polynomes quelconques*, Gauthier-Villars, Paris (1949).

LENGTHS OF PRIME IDEAL CHAINS.*

By I. S. COHEN.

Introduction. Given a commutative ring R and a prime ideal p in R, we may consider the maximal lengths of chains of prime ideals contained in p and containing p. These give rise to the *height* and *depth* of p (usually called the *rank* and *dimension*); see section 1 for the precise definitions. If S is a ring containing R, and P a prime ideal in S contracting to p, then we are concerned with the relation between the heights and depths of P and p. For Noetherian rings, with which this paper is principally concerned, partial answers are given in Theorems 1 and 4, the latter generalizing some of the results on dimension in Krull [8]. The results on height are motivated by some well known theorems on valuations of algebraic function fields; these are stated as corollaries to Theorems 1 and 3. These results will moreover be applied in a subsequent paper to the theory of valuations of local rings.

1. Elementary properties. The term *ring* will always mean a commutative ring with identity. A *proper* ideal in a ring R is an ideal different from (0) and R. The term *prime ideal* shall always refer to one which is $\neq R$, though not necessarily $\neq (0)$.

Definition. Let R be a ring. A *prime ideal chain* (or simply a *chain*) *in R of length k* is a collection of $k+1$ prime ideals in R, ordered by inclusion.

In other words, a chain of length k is a collection of prime ideals $p_0, p_1, \cdots, p_k$, which, when suitably numbered, satisfy

$$p_0 < p_1 < \cdots < p_k < R. \tag{1}$$

The chain is said to *begin* with p_0 and to *end* with p_k. The chain is said to be *maximal* if no further prime ideal can be inserted between p_{i-1} and p_i for $i = 1, 2, \cdots, k$.

* Received August 9, 1953.

[1] If A and B are sets, we write $A \subset B$ to denote that A is a subset of B and $A < B$ to denote that A is a proper subset of B. We point out explicitly that in the definition of chain the possibility $p_0 = (0)$ is *not* excluded.

Definition. Let p be a prime ideal in a ring R. The *depth* [2] (*height*) of p is the maximum length over all chains in R beginning (ending) with p.

The depth and height of p will be denoted by $d(p)$ and $h(p)$ respectively; either may be infinite. Clearly $d(p) = 0$ if and only if p is maximal; $h(p) = 0$ if and only if p contains no other prime ideal of R. It is known that in a Noetherian ring $h(p) < \infty$, but $d(p)$ may $= \infty$ [7, section 8].[3] If R is a local ring, then also $d(p) < \infty$.

If p is of height k, then there exists a chain (1) with $p = p_k$; this chain is necessarily maximal, and p_0 can contain no other prime ideal of R, so that if R is an integral domain, then $p_0 = (0)$. Not every maximal chain ending with p need be of length k, as can be shown by simple examples of rings with zero divisors. Whether this unpleasant phenomenon may occur even when R is an integral domain is an unsolved problem, which we shall refer to as "the maximal chain problem."

Still assuming that p is of height k and that $p = p_k$ in (1), we see clearly that p_i is of height i, $i = 0, 1, \cdots, k$. Also p/p_i is of height $k - i$ in the ring R/p_i. If p and q are prime ideals in R such that $q \subset p$, then clearly

$$h(q) + h(p/q) \leqq h(p). \tag{2}$$

One might hope for equality when R is an integral domain, but this question is equivalent to the maximal chain problem mentioned above.

Let M be a subset of R closed under multiplication and containing no zero divisor of R, and let R_M denote, as usual, the quotient ring. If p is a prime ideal not meeting M, then elementary properties of the quotient ring show that $h(p) = h(R_M p)$.

Observe that from (2) we may conclude that if $q < p$, then $h(q) < h(p)$.

Entirely similar statements can be made about $d(p)$.

Definition. Let R and S be rings,[4] $R \subset S$, P a prime ideal in S, $p = P \cap R$. Then the R-dimension of P, to be denoted by $\dim P/R$, is $\mathrm{dt}(S/P)/(R/p)$, where "dt" means degree of transcendence.

If R is a field, then $\dim P/R = \mathrm{dt}(S/P)/R$; this definition is well known for prime ideals in a polynomial ring over a field.

---

[2] With regard to terminology, see the remarks following Lemma 1.

[3] Numbers in brackets refer to the bibliography at the end of the paper.

[4] When we write $R \subset S$, we always assume that R and S have the same identity element.

If Q is a prime ideal in S such that $Q \subset P$, and if we place $q = Q \cap R$, then

$$\dim P/R = \dim (P/Q)/(R/q). \tag{3}$$

For inasmuch as $P/Q \cap R/q = p/q$, the right side is the degree of transcendence of $(S/Q)/(P/Q)$ over $(R/q)/(p/q)$. Since these two rings are isomorphic respectively to S/P and R/p, the right side equals $\dim P/R$.

To see the behavior of the dimension under quotient ring formation, let us assume S is an integral domain. Let M and N be multiplicatively

closed subsets of R and S respectively, not containing 0, such that $M \subset N$; then $R_M \subset S_N$. If, now, P is a prime ideal in S not meeting N, then

$$\dim (S_N P)/R_M = \dim P/R. \tag{4}$$

For inasmuch as $S_N P \cap R_M = R_M p$, the left side is the transcendence degree of $S_N/S_N P$ over $R_M/R_M p$. Now $S_N/S_N P$ can be regarded in a natural way as containing S/P, and indeed it is then a quotient ring of S/P, namely with respect to the set of the residue classes of the elements of N. Similarly, $R_M/R_M p$ can be regarded as a quotient ring of R/p. Hence the degree of transcendence just referred to equals $\mathrm{dt}(S/P)/(R/p) = \dim P/R$, and (4) is proved.

The dimension is additive—i. e., if R_0 is a subring of R, then $\dim P/R_0 = \dim P/R + \dim p/R_0$.

The elementary properties of the dimension and of the height here referred to, as well as the analogous properties of the depth, will in the sequel be freely used without further comment.

In an important elementary case, the dimension and height are closely related:

LEMMA 1. *Let R and S be integral domains, $R \subset S$, S finitely generated over R, Q a prime ideal in S such that $Q \cap R = (0)$. Then $h(Q) + \dim Q/R = \mathrm{dt}\, S/R$.*

Proof. When R is a field, this is a classical result [4, p. 43]. In the general case, let M denote the set of nonzero elements of R, so that $R_M \subset S_M$, R_M is the quotient field of R, and S_M is finitely generated over R_M. Then by

the classical result, $h(S_MQ) + \dim S_MQ/R_M = \mathrm{dt}\, S_M/R_M$. Since $\mathrm{dt}\, S_M/R_M = \mathrm{dt}\, S/R$, the lemma follows.

If R is a field, then it is also well known [loc. cit.] that $d(Q) = \dim Q/R$. For this reason the term *dimension* has generally been used for what we have called the depth. However the two are not in general equal (e. g., the maximal ideal of the local ring of a subvariety of positive dimension of an algebraic variety), and we have therefore preferred to use different terms. What we have called the height has been generally called the *rank* or the *dimension defect*, the latter term finding some justification in the formula proved in the lemma.

2. The defect. *We now introduce some notations which will be held fixed for the remainder of the paper.* R and S are integral domains such that $R \subset S$, P is prime in S,

$$p = P \cap R, \qquad \bar{R} = R/p, \qquad \bar{S} = S/P,$$

K, L, $\bar{K}$, $\bar{L}$ are the respective quotient fields of R, S, $\bar{R}$, $\bar{S}$. The capital letter X will always denote an indeterminate.

We are concerned here with the proof, under certain conditions, of the following relation:

$$h(P) + \mathrm{dt}\, \bar{S}/\bar{R} \leqq h(p) + \mathrm{dt}\, S/R. \tag{5}$$

We assume the right side to be finite since otherwise there is nothing to prove. It is then useful to study systematically the difference of the two sides of (5).

Definition. If $h(p)$ and $\mathrm{dt}\, S/R$ are finite, then the *defect of P over R*, to be denoted by $\mathrm{def}\, P/R$, is given by

$$\mathrm{def}\, P/R = \mathrm{dt}\, S/R + [h(p) - h(P)] - \dim P/R.$$

The defect of P over R is thus an integer or $-\infty$, and indeed any such value is possible for suitable R, S and P. We are to prove that under certain conditions it is non-negative.

LEMMA 2. (a) *The defect is additive. I. e., suppose R_0 is an integral domain contained in R, and let $p_0 = p \cap R$; then if $h(p_0)$, $h(p)$, and $\mathrm{dt}\, S/R_0$ are finite, then* $\mathrm{def}\, P/R_0 = \mathrm{def}\, P/R + \mathrm{def}\, p/R_0$.

(b) *Let M and N be multiplicatively closed subsets of R and S respectively such that $M \subset N$ and N does not meet P. If either one of* $\mathrm{def}\, P/R$, $\mathrm{def}\, S_NP/R_M$ *is defined, then so is the other, and they are equal.*

(c) *Let Q be a prime ideal of S contained in P, let $Q \cap R = (0)$, and let S be finitely generated over R. If $h(p)$ is finite, then* $\operatorname{def} P/R \leqq \operatorname{def}(P/Q)/R$, *with equality holding if and only if* $h(P) = h(Q) + h(P/Q)$.

(d) *If S is integrally dependent on R and $h(p)$ is finite, then* $\operatorname{def} P/R \geqq 0$; *and if R is integrally closed, then* $\operatorname{def} P/R = 0$.

Proof. (a) It follows from the hypothesis of (a) that all three defects are defined. The additivity then follows from the fact that all three terms of which the defect is a sum are additive.

(b) This follows from facts previously stated about dimension and height.

(c) Note that R can be imbedded in natural fashion into S/Q. When this is done, then $(P/Q) \cap R = p$. Now $\dim P/R = \dim (P/Q)/R$. From Lemma 1,

$$\begin{aligned} \operatorname{dt} S/R - h(P) &= h(Q) + \dim Q/R - h(P) \\ &= \operatorname{dt}(S/Q)/R + [h(Q) - h(P)] \leqq \operatorname{dt}(S/Q)/R - h(P/Q), \end{aligned}$$

the last inequality because $h(P) \geqq h(Q) + h(P/Q)$. Thus (c) is proved.

(d) In this case, $\operatorname{dt} S/R = \operatorname{dt} \bar{S}/\bar{R} = 0$. That $h(P) \leqq h(p)$ follows from the fact that the contraction to R of any chain ending in P yields a chain ending in p [5, Theorem 2], [1, Theorem 4]. If R is integrally closed, then any chain of, say, m prime ideals ending in p gives rise to a chain of length m ending in P. Namely, take $P_m = P$. By the "going down" theorem [5, Theorem 6], [1, Theorem 5], P_m contains a prime ideal P_{m-1} lying over p_{m-1}; and we may obviously continue in this fashion. Both parts of (d) were proved by Krull [5, Theorems 7, 8]. It seems to be still an unsolved question whether inequality may actually hold when R is not integrally closed.

Before proceeding to the proof of (5), we prove two well known lemmas on polynomial rings.

LEMMA 3. *Let $R' = R[X]$, let p'_1 and p'_2 be two prime ideals in R' having the same contraction p in R, with $p'_1 < p'_2$. Then $p'_1 = R'p$.*

Proof. We note that $R'/R'p$ is a polynomial ring over R/p. In $R'/R'p$, $p'_1/R'p$ and $p'_2/R'p$ are prime ideals contracting in R/p to (0). Since we are trying to prove that $p'_1/R'p = (0)$, it follows that it is sufficient to prove the lemma under the additional hypothesis that $p = (0)$. If K is the

quotient field of R, then $K[X]$ is the quotient ring of $R[X]$ with respect to the multiplicatively closed set of nonzero elements of R. Since neither p'_1 nor p'_2 meets this set, their extensions to $K[X]$ are prime ideals, with $K[X]p'_1 < K[X]p'_2 < K[X]$. Since K is a field, this requires $K[X]p'_1 = (0)$.

COROLLARY. *If $p'_1 < p'_2 < p'_3$ are prime ideals in $R[X]$, they cannot all have the same contraction in R.*

LEMMA 4. *Let R be a Noetherian integral domain, $R' = R[X]$. If p is a minimal prime ideal in R (i. e., the only prime ideal properly contained in p is (0)), then $R'p$ is minimal in R'.*

Proof. If $y \in p$, $y \neq 0$, then p is an isolated prime ideal belonging to Ry, hence $R'p$ is an isolated prime ideal belonging to $R'y$. Since R' is Noetherian, the principal ideal theorem [7, Theorem 7*] implies that $R'p$ is minimal in R'.

3. The main theorems on height. We are now in a position to prove the basic inequality (5) for the case where R is Noetherian. In order better to see just where the Noetherian character comes in, we state our theorem for a (potentially) wider class of rings. Whether this class is actually larger will not be discussed in the present paper.

THEOREM 1. *Let $\mathfrak{D}$ be a collection of integral domains having the following properties*:

(a) *If $A \in \mathfrak{D}$ and if q is a prime ideal in R, then $A/q \in \mathfrak{D}$.*

(b) *If $A \in \mathfrak{D}$, then $A[X] \in \mathfrak{D}$.*

(c) *If $A \in \mathfrak{D}$ and if q is a minimal prime ideal in A, then $A[X]q$ is a minimal prime in $A[X]$.*

Under these conditions, if $R \in \mathfrak{D}$, then the inequality (5) holds.[5]

Remarks. The collection of all Noetherian integral domains satisfies (a), (b), (c), the last by Lemma 4. If $A \in \mathfrak{D}$ then (b) implies that any polynomial ring over A in a finite number of indeterminates is also in $\mathfrak{D}$, and hence by (a) any integral domain finitely generated over A is in $\mathfrak{D}$. Observe that it is not required that S be in $\mathfrak{D}$.

Proof. We assume that $h(p)$ and $\operatorname{dt} S/R$ are finite and shall show that $\operatorname{def} P/R \geqq 0$.

[5] We point out once more that the notation introduced at the beginning of section 2 is to be in force for the remainder of the paper.

Let m be any positive integer $\leqq h(P)$ (which may a priori be infinite). Then there exists in S a chain

$$P_0 < P_1 < \cdots < P_m = P. \tag{6}$$

For $i = 1, 2, \cdots, m$, let $x_i \in P_i$, $x_i \notin P_{i-1}$. If n is any non-negative integer $\leqq \operatorname{dt} \bar{S}/\bar{R}$, then there exist in $\bar{S}$ elements $\bar{y}_1, \cdots, \bar{y}_n$ algebraically independent over $\bar{R}$. Let y_j be a representative of $\bar{y}_j$ in S, and place

$$R' = R[x_1, \cdots, x_m; y_1, \cdots, y_n],$$

$$p' = P \cap R', \qquad p'_i = P_i \cap R', \qquad \bar{R}' = R'/p'.$$

Thus

$$R \subset R' \subset S, \qquad p' \cap R = p, \qquad p'_0 < p'_1 < \cdots < p'_m = p',$$
$$h(p') \geqq m, \operatorname{dt} \bar{R}'/\bar{R} \geqq n.$$

Assuming for the moment that we have proved that $\operatorname{def} p'/R \geqq 0$, we have $h(p') + \operatorname{dt} \bar{R}'/\bar{R} \leqq h(p) + \operatorname{dt} R'/R$, hence $m + n \leqq h(p) + \operatorname{dt} S/R$ for all $m \leqq h(P)$ and all $n \leqq \operatorname{dt} \bar{S}/\bar{R}$. And this proves the theorem. So it is enough to prove that $\operatorname{def} p'/R \geqq 0$, and since R' is finitely generated over R we may henceforth confine ourselves to proving the theorem in the case where S is finitely generated over R.

Consider first the special case where S is a polynomial ring in one variable, $S = R[X]$. We proceed by induction on $h(p)$. If $h(p) = 0$, then $p = (0)$, and we must show that $h(P) + \operatorname{dt} \bar{S}/\bar{R} \leqq 1$. If $P = (0)$, this is trivial; and if $P \neq (0)$ then $\bar{S}$ is algebraic over $\bar{R}$, and by Lemma 3, Cor., P is minimal, so that $h(P) = 1$.

Now suppose $h = h(p) > 0$. If m is any positive integer $\leqq h(P)$, then there exists in S a chain of the type (6) above. It is enough to show that $m + \operatorname{dt} \bar{S}/\bar{R} \leqq h + 1$. If $m = 1$, this is clear, so assume $m \geqq 2$. Again by the corollary to Lemma 3, $P_2 \cap R$ is a nonzero ideal, necessarily of finite height. So it contains a minimal prime ideal p_0, and obviously $p_0 \subset p$. By hypothesis (c), Sp_0 is minimal in S, hence $Sp_0 \neq P_2$. Passing to S/Sp_0 and denoting images in this ring by an asterisk, we have $h(P^*) \geqq m - 1$. The ideal $p^* = p/p_0$ is prime in $R^* = R/p_0$ and clearly $h(p^*) \leqq h - 1$, $P^* \cap R^* = p^*$. Moreover S^* is a polynomial ring over R^* and $R^* \in \mathfrak{D}$ by hypothesis (a). Hence the induction hypothesis gives $h(P^*) + \dim P^*/R^* \leqq h(p^*) + 1$. But $\dim P^*/R^* = \dim P/R$, so

$$m - 1 + \dim P/R \leqq (h - 1) + 1,$$

and this completes the proof for the case $S = R[X]$.

By induction it now follows for S a polynomial ring in any finite number of variables in view of the additivity of the defect. Induction is permitted by hypothesis (b).

Finally, let S be any integral domain finitely generated over R. Then S is the homomorphic image in a natural way of a polynomial ring S_1 over R. Let Q_1 be the kernel and let P_1 be the complete inverse image of P. Then

$$Q_1 \subset P_1, \qquad Q_1 \cap R = (0), \qquad P_1 \cap R = p, \qquad P_1/Q_1 = P.$$

By the polynomial case, def $P_1/R \geqq 0$. By Lemma 2(c), we have def $P/R \geqq$ def P_1/R. The proof is thus complete.

Lemma 2(c), which we have just used, depends on Lemma 1, which in turn depends on the dimension theory of polynomial rings over a field. We could avoid the use of this dimension theory as follows. To prove Theorem 1 for S finitely generated over R, we may assume $S = R[x]$, since additivity of the defect permits induction. If x is transcendental we proceed as in the proof above. If x is algebraic, then S is the homomorphic image of a polynomial ring $S_1 = R[X]$, in *one* variable. With the notation of the preceding paragraph, we have that $h(P_1) + \dim P_1/R \leqq h(p) + 1$. But $h(P_1) \geqq h(P) + 1$ and $\dim P_1/R = \dim P/R$, whence the desired result. We have, in effect, used the fact that Lemma 1 is trivial if $S = R[X]$ and Q is minimal.

COROLLARY 1. *Let K be an algebraic function field of n variables over a field F. If v is a valuation of K over F of rank m and dimension d, then $m + d \leqq n$.*

Proof. If S is the valuation ring of v and P is the maximal ideal of S, then it is well known that $h(P) = m$. Taking $R = F$, we get the stated result, since by definition, $d = \mathrm{dt}\, \bar{S}/F$.

COROLLARY 2. *If $R \in \mathfrak{D}$ and S is a polynomial ring over R in a finite number of indeterminates, then equality holds in* (5).

Proof. We may assume one indeterminate, $S = R[X]$, $\mathrm{dt}\, S/R = 1$. If $P = Sp$, then $\mathrm{dt}\, \bar{S}/\bar{R} = 1$, so the theorem reads $h(P) \leqq h(p)$. The opposite inequality is obtained by extending to S any chain in R ending in p. If $Sp < P$, then $\mathrm{dt}\, \bar{S}/\bar{R} = 0$, so the theorem reads $h(P) \leqq h(p) + 1$. But also $h(P) \geqq h(Sp) + 1 = h(p) + 1$.

One might attempt to extend Cor. 2 to the case where S is any domain finitely generated over R. If we proceed as in the last paragraph of the

proof of the theorem, success would require that we have def $P/R = \text{def } P_1/R$. Referring to the proof of Lemma 2(c), this equality is equivalent to $h(P) = h(Q) + h(P/Q)$ (in the notation of that lemma). So this extension of Cor. 2 would be equivalent to a special case of the maximal chain problem. Even this special case appears to be unsolved.

But it is easy to see that the conclusion of Cor. 2 may fail if S is not finitely generated over R. For let R be a polynomial ring over a field F in n variables $X_1, \cdots, X_n$, let $p = (X_1, \cdots, X_n)$. Let S be the valuation ring of a rank 1, 0-dimensional valuation of $F(X_1, \cdots, X_n)$ whose center in R is at p. If P is the maximal ideal of S, then $\text{dt } S/R = \text{dt } \bar{S}/\bar{R} = 0$, $h(P) = 1$, $h(p) = n$, $\text{def } P/R = n - 1$. This shows, by the way, that the defect may be any non-negative integer.

Condition (c) of Theorem 1 asserts, in effect, that the conclusion of the theorem is true in a special case, namely S a polynomial ring over R in one variable and $h(p) = 1$. So we are able to conclude inequality (5) in general from its validity in a special case. Along the same lines we have

THEOREM 2. *Let $\mathfrak{D}$ be a collection of integral domains satisfying conditions* (a) *and* (b) *of Theorem* 1, *as well as*

(c′) *If $R \in \mathfrak{D}$, if p is minimal in R, and if S has the same quotient field as R, then* $\text{def } P/R \geqq 0$.

Under these conditions, inequality (5) *holds generally, provided only that $R \in \mathfrak{D}$.*

Proof. It is necessary only to prove that (c) of Theorem 1 holds. So let $A \in \mathfrak{D}$, q be minimal in A, and suppose $A[X]\,q$ is not minimal, so that there exists a prime ideal Q such that $(0) < Q < A[X]q$. Let $S = A[X]/Q$, and let P denote the image of $A[X]q$ in this ring. By virtue of Lemma 3, there is no prime ideal between Q and $A[X]q$, hence $h(P) = 1$. Since $Q \cap A = (0)$, S contains A and is algebraic over it. Also $\dim P/A = \dim A[X]q/A = 1$, so $\text{def } P/A = -1$. But this easily leads to a contradiction. For let x be the residue class of X in S. Let a be the leading coefficient of a nonzero polynomial in Q, so ax is integrally dependent on A. Let $R = A[ax]$, so $R \in \mathfrak{D}$, and $S = R[x]$ has same quotient field as R. Let $p = P \cap R$, so $p \cap A = q$, and hence p is minimal [1, Theorem 4]. By (c′), $\text{def } P/R \geqq 0$. Since R is integrally dependent on A, $\text{def } p/A \geqq 0$ by Lemma 2(d). Hence $\text{def } P/A \geqq 0$, contradiction.

4. Additional results on height. We now give an example to show that the conclusion of Theorem 1 is not true for an arbitrary domain R.

For suppose it were and suppose R had a single proper prime ideal p. Then every valuation ring of the quotient field of R which contains R would have to be of rank 1. So if R were integrally closed, so that it was the intersection of valuation rings S, then all of these would be of rank 1, hence R would be fully integrally closed (i. e., if $a \in K$ and if $b \in R$, $b \neq 0$, and $ba^n \in R$ for all n, then $a \in R$). So we need an integrally closed ring with a single proper prime ideal and which is not fully integrally closed. A simple example of such a ring has been given by Krull [6, p. 670].

The construction of this ring is as follows. Let F be a field, $K = F(X, Y)$ a rational function field in two variables. Let R consist of all rational functions $f(X, Y)$ such that $f(0, Y)$ is finite and in F. The set of all $f \in R$ such that $f(0, Y) = 0$ is clearly a maximal ideal p containing all non-units, and $R/p \cong F$. Moreover p is minimal—indeed every proper ideal A is primary for p. For let $g(X, Y)$ be a nonzero element of A and write $g(X, Y) = X^r g_1(X, Y)$, with $g_1(0, Y) \neq 0$; if $f(X, Y) \in p$, then $f(X, Y) = Xf_1(X, Y)$ with $f_1(0, Y)$ finite, so $f^{r+1}/g = Xf_1{}^{r+1}/g_1 \in p \subset R$, hence $f^{r+1} \in Rg \subset A$. If $f(X, Y) \in K$ and is integrally dependent on R, then $f(0, Y)$ is finite; moreover $f(0, Y)$ would be integrally dependent on F, hence in F, so $f(X, Y) \in R$. Thus R is integrally closed. But not fully, for $Y \notin R$, whereas $XY^n \in R$ for all n.

An explicit case of a P with $\operatorname{def} P/R < 0$ is obtained by taking S to be the set of all $f(X, Y)$ with $f(0, Y) = 0$, or else $f(0, Y)$ finite and $\neq 0$, and $f(0, 0)$ finite. This is the valuation ring for the rank 2 valuation of K obtained by composing the valuation of $F(X, Y)$ defined by the Y-axis with the valuation of $F(Y)$ defined by the origin. If P is the maximal ideal of S, then $S/P = F$, so $\operatorname{dt} \bar{S}/\bar{R} = 0$; since $h(P) = 2$, $h(p) = 1$, we have $\operatorname{def} P/R = -1$.

THEOREM 3. *Let R be Noetherian and let L be finitely generated over K. If $\operatorname{def} P/R = 0$, then $\bar{L}$ is finitely generated over $\bar{K}$. Furthermore, if $m = h(P)$ and if*

$$(0) = P_0 < P_1 < \cdots < P_m = P \tag{6}$$

is any chain of length m ending in P, then the ascending chain condition holds for those ideals between P_{i-1} and P_i which are primary for P_i $(i = 1, 2, \cdots, m)$.

Proof. Observe that $h(p)$ is finite since R is Noetherian; also $\operatorname{dt} S/R = \operatorname{dt} L/K$. Hence $\operatorname{def} P/R$ is indeed defined.

Let $x_i \in P_i$, $x_i \notin P_{i-1}$, $i = 1, 2, \cdots, m$ (cf. proof of Theorem 1). If $n = \text{dt } \bar{S}/\bar{R}$, let $\bar{y}_1, \cdots, \bar{y}_n$ be elements of $\bar{S}$ algebraically independent over $\bar{R}$. let y_j be a representative of $\bar{y}_j$ in $\bar{S}$. Since L is finitely generated over K, we may write $L = K(z_1, \cdots, z_r)$, $z_k \in S$. Now place

$$R' = R[x_1, \cdots, x_m; y_1, \cdots, y_n; z_1, \cdots, z_r],$$

$p' = P \cap R'$, $p'_i = P_i \cap R'$, $\bar{R}' = R'/p'$, $\bar{K}' =$ quotient field of $\bar{R}'$. Now $\bar{R}'$ is clearly finitely generated over $\bar{R}$, hence likewise $\bar{K}'$ over $\bar{K}$, so it is enough to prove that $\bar{L}$ is finitely generated over $\bar{K}'$. In other words, it is enough to prove the theorem for S and R'. These two rings satisfy the hypotheses of our theorem: R' is surely Noetherian. From $0 = \text{def } P/R = \text{def } P/R' + \text{def } p'/R$, we get $\text{def } P/R' = 0$. S and R' both have L as quotient field since $z_k \in R'$. Also $\text{dt } \bar{S}/\bar{R}' = 0$, since $\bar{R}' \supset \bar{R}[\bar{y}_1, \cdots, \bar{y}_n]$. Hence $h(p') = h(P) = m$. Finally, we have $(0) = p'_0 < p'_1 < \cdots < p'_m = p'$, since $x_i \in p'_i$, $x_i \notin p'_{i-1}$.

Writing R for R', we are reduced to proving our theorem under the following additional hypotheses: $\text{dt } S/R = 0$, $\text{dt } \bar{S}/\bar{R} = 0$, the contractions $P_i \cap R$ are distinct (the P_i being as in (6)). We then have $h(p) = h(P) = m$, and $(0) = p_0 < p_1 < \cdots < p_m = p$, where $p_i = P_i \cap R$.

We prove the theorem under these additional hypotheses by inductions on m. For $m = 0$, the theorem is trivial. Now assume the theorem has been proved for height $m - 1$. Let R' and S' be the quotient rings of R and S with respect to p_1 and P_1 respectively. Since p_1 is a minimal prime ideal, R' has only one proper prime ideal, namely $R'p_1$, hence R' satisfies the restricted minimum condition [2, Theorem 1]. From [2, Theorem 3], it now follows that $S'/S'P_1$ is finite-dimensional over $R'/R'p_1$, and that the ascending chain condition holds in S'.[6] The latter implies the conclusion of the theorem for the primary ideals belonging to P_1.

Passing to the rings S/P_1 and R/p_1, we have the ideals P/P_1 and p/p_1 both clearly of height $m - 1$. The quotient fields of these rings are $S'/S'P_1$ and $R'/R'p_1$. Furthermore $(S/P_1)/(P/P_1)$ is algebraic over $(R/p_1)/(p/p_1)$, since these are isomorphic respectively to $\bar{S}$ and $\bar{R}$. From all this, it follows that S/P_1, R/p_1, and P/P_1 satisfy not only the hypotheses of our theorem, but also the additional hypotheses stated in the second preceding paragraph. By induction hypothesis, $\bar{L}$ is finitely generated over $\bar{K}$. For the same reason the ascending chain condition holds for the ideals between P_{i-1}/P_1 and P_i/P_1

[6] The quoted theorem is proved under the assumption that S is integral over R. But a trifling modification in the proof suffices to cover the case where S is merely algebraic over R.

which are primary for P_i/P_1 $(i = 2, \cdots, m)$. This implies the desired result for the ideals between P_{i-1} and P_i.

From the theorem we can derive a well known result on valuations.

COROLLARY. *Let K be an algebraic function field of n variables over a field F. If v is a valuation of K over F of rank m and dimension d, and if $m + d = n$, then v is discrete and its residue field is finitely generated over F.*

Proof. (Cf. Theorem 1, Cor. 1.) Take S to be the valuation ring of v, P its maximal ideal, $R = F$. S contains a unique chain of type (6) above. The theorem states that the quotient ring of S/P_{i-1} with respect to P_i/P_{i-1} is Noetherian, $i = 1, 2, \cdots, m$. So each of the m rank 1 valuations of which v is composed is discrete, so v is discrete.

5. Depth of prime ideals. We retain in this section, as throughout, the notation introduced in section 2.

We shall obtain here Krull's results on depth of prime ideals [8, Theorems 13, 14, 16, 18] in somewhat stronger form, and with rather a simpler proof.

LEMMA 5. *Let R be any ring, $R' = R[X]$. If q is prime in R, then $d(R'q) \geqq d(q) + 1$.*

Proof. If $m \leqq d(q)$, then there exists a chain $q = q_0 < q_1 < \cdots < q_m$ in R. Then a chain in R' is given by

$$R'q = R'q_0 < R'q_1 < \cdots < R'q_m < R'q_m + R'X,$$

so that $d(R'q) \geqq m + 1$, whence the lemma.

THEOREM 4. *Let $\mathfrak{D}$ be a collection of integral domains satisfying conditions* (a), (b), (c) *of Theorem* 1. *Then, if $R \in D$,*

$$d(P) \leqq d(p) + \dim P/R. \tag{7}$$

If S is finitely generated over R, and if R is Noetherian, then the right side exceeds the left by at most unity. If, in addition, R is a Hilbert[7] ring, then equality holds.

Proof. Since the quantities involved in (7) are properties of the residue class rings S/P and R/p, we may assume without loss of generality that $P = (0)$. So we must prove that

$$d(S \cdot 0) \leqq d(R \cdot 0) + \mathrm{dt}\, S/R. \tag{8}$$

[7] A *Hilbert ring* is a ring in which every prime ideal is the intersection of maximal ideals [3]. Krull [8] uses the term *Jacobson ring*.

Now if P_1 is prime in S, then by our main result on height (Theorem 1),

$$h(P_1) \leqq h(P_1 \cap R) + \text{dt } S/R \leqq d(R \cdot 0) + \text{dt } S/R.$$

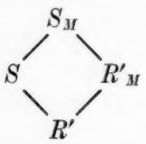
Since $d(S \cdot 0)$ is the least upper bound of all $h(P_1)$, (8) follows. It is easy to avoid explicit use of the theorem on heights; for this see the remarks at the end of the proof.

Now let S be finitely generated over R and R be Noetherian (note that our reduction to the case $P = (0)$ has not disturbed these conditions). Then we must show that $d(S \cdot 0) \geqq d(R \cdot 0) + \text{dt } S/R - 1$. Let $t = \text{dt } S/R$, let $x_1, \cdots, x_t$ be a transcendence basis for S over R, let $R' = R[x_1, \cdots, x_t]$. By Lemma 5, $d(R' \cdot 0) \geqq d(R \cdot 0) + t$ (indeed, equality holds by virtue of the first part of the theorem); hence we must prove that

$$d(S \cdot 0) \geqq d(R' \cdot 0) - 1, \tag{9}$$

this with the information that S is finitely generated and algebraic over R'. So let $S = R'[y_1, \cdots, y_m]$. Since y_i is algebraic over R', it is a zero of a certain polynomial in $R'[X]$; we may assume that each polynomial has the same leading coefficient a, and of course $a \neq 0$. If M is the multiplicative system consisting of the non-negative powers of a, then each y_i is integral over R'_M, hence S_M, which equals $R'_M[y_1, \cdots, y_m]$, is integral over R'_M.

If d is any integer $\leqq d(R' \cdot 0)$, then we show that $d(S \cdot 0) \geqq d - 1$.



We have a chain

$$R' \cdot 0 < p_1 < p_2 < \cdots < p_{d-1} < p_d. \tag{10}$$

We claim that the chain may be selected so that $a \notin p_i$, $i = 1, 2, \cdots, d-1$. This is an old lemma of Krull, but we sketch a proof by induction on d. Namely let $q_1, q_2, \cdots$ be the minimal prime divisors of $R' \cdot a$; they are minimal in R', hence cannot contain p_2. So there exists an element b in p_2 not in any of the q_i; p_2 must contain a minimal prime divisor q of $R' \cdot b$, $q < p_2$, $a \notin q$. In R'/q, we have a chain $(0) < p_2/q < \cdots < p_d/q$, and the residue class of a is not zero. Applying induction hypothesis here and mapping back to R', we find our claim substantiated.

Thus in (10), we assume $a \notin p_i$, $i = 1, \cdots, d-1$. By the elementary properties of quotient rings we have in R_M a chain

$$R'_M \cdot 0 < R'_M p_1 < R'_M p_2 < \cdots < R'_M p_{d-1}. \tag{11}$$

Since S_M is integrally dependent on R'_M, the "going up" theorem [1, Theorem 3] gives us in S_M a chain $S_M \cdot 0 < P'_1 < \cdots < P'_{d-1}$ lying over the chain (11). Contracting to S, we get a chain

$$S \cdot 0 < P_1 < \cdots < P_{d-1}, \tag{12}$$

and clearly $P_i \cap R' = p_i$, $i = 1, \cdots, d-1$. Hence $d(S \cdot 0) \geqq d-1$.

Now if R were originally a Hilbert ring, then it remains so after the preliminary passage to the case $p = (0)$ [3, Theorem 2], and then R' is likewise a Hilbert ring [3, Theorem 3]. Since S is finitely generated over R', and R' is Hilbert, every maximal ideal in S contracts to a maximal ideal in R' [3, Theorem 5].[8] It follows that in the prime ideal chain (12), P_{d-1} is not maximal, hence $d(S \cdot 0) \geqq d$. If $R' \cdot 0$ has finite depth, we take $d = d(R' \cdot 0)$ and have $d(S \cdot 0) \geqq d(R' \cdot 0) \geqq d(R \cdot 0) + t$, as was to be shown. This completes the proof of Theorem 4.

If the assumption that R is Hilbert is omitted, then proper inequality may well hold in (7), even though R is Noetherian [8, Theorem 4]. For there will exist in $R[X]$ a maximal ideal P whose contraction to R is a non-maximal ideal p [3, Theorem 5]. Thus $d(P) = 0$, $d(p) > 0$.

As indicated above, the proof of inequality (7) can be carried through without specific appeal to Theorem 1. First:

LEMMA 6. *Let $\mathfrak{D}$ be as in Theorem 4, $R \in \mathfrak{D}$, $R' = R[X]$. If p_1 and p_2 are two prime ideals in R such that $p_1 < p_2$ and there are no intermediate prime ideals, then there are likewise none between $R'p_1$ and $R'p_2$.*

Proof. Since $R'/R'p_1$ is a polynomial ring over R/p_1, it is sufficient to prove the lemma in the case $p_1 = (0)$. But then it follows from condition (c) on $\mathfrak{D}$.

We now prove (7), first for the case that S is a polynomial ring over R in one variable and $P = Sp$; we may assume $d(p)$ is finite. Then $\dim P/R = 1$ and we must show that $d(P) \leqq d(p) + 1$. If $d(p) = 0$, then p is maximal, S/Sp is a polynomial ring over the field R/p, so $d(Sp) = 1$. We then proceed by induction on $d(p)$. Suppose that we have in S a chain $Sp < P_1 < P_2 < \cdots < P_r$; we shall prove that $r \leqq d(p) + 1$. By Lemma 3, Cor., $P_2 \cap R$ cannot equal p, so it contains some prime ideal p' such that $p < p'$ and such that there is no prime ideal between p and p'. By Lemma 6, there is none between Sp and Sp', so $Sp' \neq P_2$. The chain $Sp' < P_2 < \cdots$

[8] The quoted theorem refers to polynomial rings, but the result is then obviously valid also for homomorphic images of such rings.

$< P_r$ shows that $d(Sp') \geqq r-1$; and obviously $d(p') \leqq d(p) - 1$. But the induction hypothesis shows that $d(Sp') \leqq d(p') + 1$, whence $r \leqq d(p) + 1$.

Still continuing with the case $S = R[X]$, suppose $P > Sp$. Then $\dim P/R = 0$, $d(P) < d(Sp) \leqq d(p) + 1$, $d(P) \leqq d(p)$.

Induction now takes care of the case where S is a polynomial ring over R in several variables, and residue class ring formation suffices to handle the case S finitely generated over R.

To conclude the proof for the most general case, we once more reduce to the case $P = (0)$. Let m be any positive integer $\leqq d(S \cdot 0)$ (which may be infinite). Then there exists a chain $S \cdot 0 = P_0 < P_1 < P_2 < \cdots < P_m$. Let $x_i \in P_i$, $x_i \notin P_{i-1}$, $i = 1, \cdots, m$ (cf. proof of Theorem 1), and let $R' = R[x_1, \cdots, x_m]$. Since R' is finitely generated over R, we have $d(R' \cdot 0) \leqq d(R \cdot 0) + \operatorname{dt} R'/R$. But clearly $m \leqq d(R' \cdot 0)$, hence

$$m \leqq d(R \cdot 0) = \operatorname{dt} S/R$$

for all $m \leqq d(S \cdot 0)$. Thus (7) is proved for $S \cdot 0$.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

BIBLIOGRAPHY.

[1] I. S. Cohen and A. Seidenberg, "Prime ideals and integral dependence," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 252-261.

[2] I. S. Cohen, "Commutative rings with restricted minimum condition," *Duke Mathematical Journal*, vol. 17 (1950), pp. 27-42.

[3] Oscar Goldman, "Hilbert rings and the Hilbert Nullstellensatz," *Mathematische Zeitschrift*, vol. 54 (1951), pp. 136-140.

[4] W. Krull, *Idealtheorie*, Ergebnisse der Mathematik und Ihrer Grenzgebiete, vol. 4, no. 4, Berlin, 1935.

[5] ———, "Zum Dimensionsbegriff der Idealtheorie," *Mathematische Zeitschrift*, vol. 42 (1937), pp. 745-766.

[6] ———, "v-Ideale und vollständig ganz abgeschlossene Integritätsbereiche," *Mathematische Zeitschrift*, vol. 41 (1936), pp. 665-679.

[7] ———, "Dimensionstheorie in Stellenringen," *Journal für die Reine und Angewandte Mathematik*, vol. 179 (1938), pp. 204-226.

[8] ———, "Jacobsonsche Ringe, Hilbertscher Nullstellensatz, Dimensionstheorie," *Mathematische Zeitschrift*, vol. 54 (1951), pp. 354-387.

ON THE STRUCTURE OF A CERTAIN CLASS OF KAEHLER VARIETIES.* [1]

By JUN-ICHI IGUSA.

It is a classical theorem that simply connected Riemann surfaces can be divided into three types. In the case of several complex variables the circumstance is not so simple, but there is a generalization in the following sense. Let V^n be a compact complex variety carrying a Kählerian metric with constant curvature; then its universal covering variety $\bar{V}^n$ has only three complex structures. In fact $\bar{V}^n$ is the complex projective space S^n, the interior of a unit sphere K^n or the complex vector space C^n according as the curvature is positive, negative or zero. They are known respectively as elliptic, hyperbolic or parabolic cases. We conclude that V^n itself coincides with S^n in the elliptic case, which gives a characterization of the complex projective space in terms of global differential geometry. Moreover in the hyperbolic case V^n can be considered as a non-singular projective model. Finally in the parabolic case V^n has a finite covering variety of complex torus T^n such that the types of extensions are finite for each n. We note also that V^n are minimal models in the hyperbolic and parabolic cases. The "minimality" seems to be closely connected with the "uniformizability" of compact complex varieties. Finally we shall give an example of parabolic variety with the first irregularity equal to zero, which can only happen for $n \geqq 3$.

1. Local consideration. Let $g_{\alpha\beta^*}(z, \bar{z})$ $(1 \leqq \alpha, \beta \leqq n)$ be n^2 power series in $2n$ variables z_α, $\bar{z}_\beta$ $(1 \leqq \alpha, \beta \leqq n)$ forming a positive definite Hermitian matrix in some neighborhood of the origin. In this case we can take $ds^2 = 2 \cdot \sum_{\alpha,\beta} g_{\alpha\beta^*}(z, \bar{z})\,(dz^\alpha d\bar{z}^\beta)$ as a metric ground form in local Hermitian geometry. We shall assume thereby that this metric is *without torsion*, or, in an apparently stronger form ([7], pp. 175-176), that $g_{\alpha\beta^*}(z, \bar{z})$ can be derived from a local potential $U(z, \bar{z})$ in the form

$$g_{\alpha\beta^*}(z, \bar{z}) = \partial^2 U(z, \bar{z})/\partial z^\alpha \partial \bar{z}^\beta \qquad (1 \leqq \alpha, \beta \leqq n).$$

* Received May 25, 1953; revised February 11, 1954.

[1] This work was partially supported by a research project at Harvard University, sponsored by the Office of Ordnance Research, U. S. Army.

We say that the metric has a *constant curvature* when Riemann's curvature tensor in our geometry satisfies the relation

$$R_{\alpha\beta^*\gamma\delta^*} + k \cdot (g_{\alpha\beta^*}g_{\gamma\delta^*} + g_{\gamma\beta^*}g_{\alpha\delta^*}) = 0 \qquad (1 \leqq \alpha, \beta, \gamma, \delta \leqq n)$$

with some constant k. This definition is a natural counterpart of the real case (cf. [2], Theorem 5) and the following theorem is proved by Bochner ([2], Theorem 6).

LEMMA 1. *There exists an analytic coordinate transformation after which the local potential has the normal form*

$$U(z, \bar{z}) = \begin{cases} (1/k) \cdot \log(1 + k \cdot \sum_{\alpha} z_\alpha \bar{z}_\alpha) & (k \neq 0) \\ \sum_{\alpha} z_\alpha \bar{z}_\alpha & (k = 0). \end{cases}$$

The metric is called *elliptic, hyperbolic* or *parabolic* according as k is positive, negative or zero. If we multiply the absolute value of k to our metric, we may assume that k is ± 1 in the case $k \neq 0$. We shall endow elliptic, hyperbolic and parabolic metrics to S^n, K^n and C^n globally and we call them *normal domains.* The normal domain $\mathbf{D}$ admits a well-known group $\mathbf{G}$ of analytic isometries, which is in any case a *transitive* Lie group of $n^2 + 2n$ real parameters. Thus $\mathbf{D}$ is a homogeneous space with the "structure group" $\mathbf{G}$ and with the unitary group $U(n)$ as its "isotropy group." Our later consideration is based on the following remark.

LEMMA 2. *Let Q_1 and Q_2 be two points in the normal domain $\mathbf{D}$ and let σ be an analytic local isometry sending Q_1 to Q_2, then σ can be extended uniquely to some global isometry of $\mathbf{D}$ belonging to $\mathbf{G}$.*

Proof. Since $\mathbf{G}$ is transitive on $\mathbf{D}$, we have only to prove the following assertion. Let $f_\alpha(z)$ $(1 \leqq \alpha \leqq n)$ be n analytic functions defining an isometry between some neighborhoods of the origin, then it holds $f_\alpha(z) = \sum_{\beta} a_{\alpha\beta} z_\beta$ $(1 \leqq \alpha \leqq n)$ with $(a_{\alpha\beta})$ in $U(n)$. In fact let Σ be a sufficiently small geodesic sphere about the origin, then the image Σ^* of Σ is also a geodesic sphere with the same radius about the origin. However Σ and Σ^* are ordinary spheres, hence our lemma follows from a result of Bochner ([3], pp. 58-59).

We conclude from this lemma that $\mathbf{G}$ is the full group of analytic isometries of $\mathbf{D}$. Moreover since $\mathbf{D}$ admits only one invariant metric with respect to $\mathbf{G}$, the hyperbolic metric is also Bergmann's metric (cf. [3], pp. 121-122). Hence $\mathbf{G}$ is the full group of analytic homeomorphisms of $\mathbf{D}$ in this case.

2. Global consideration. Let $\bar{V}^n$ be an n-dimensional complex variety which is *simply connected.* A global Hermitian metric on $\bar{V}$ is called a *Kählerian metric* when it is everywhere without torsion. We shall assume that $\bar{V}$ has a Kählerian metric with constant curvature, then we can prove the following

LEMMA 3. *There exists a single-valued analytic isometry* Ψ *from* $\bar{V}$ *into the normal domain* **D**.

Proof. Let P_0 be any but fixed point of $\bar{V}$ and let ψ_0 be a "germ" of an analytic isometry from $\bar{V}$ into **D** sending P_0 to the origin of **D**. The existence of ψ_0 is assured by Lemma 1. We then take an arbitrary point P_1 of $\bar{V}$ and connect it to P_0 by a continuous path P_t $(0 \leqq t \leqq 1)$. In the following we shall show that there exists one and only one family ψ_t $(0 \leqq t \leqq 1)$ of germs of analytic isometries from $\bar{V}$ into **D** sending P_t to $\psi_t(P_t)$, which is "coherent" in the sense that ψ_t induces $\psi_{t'}$ when $P_{t'}$ is near enough to P_t. Since the uniqueness is evident, we shall prove the existence and that we have only to prove the following assertion. Let ψ_t $(0 \leqq t < 1)$ be a family of coherent isometries which is open on the right side, then we can complete it by ψ_1.

Let ϕ be an analytic isometry of some neighborhood U of P_1 into **D** sending P_1 to the origin of **D**. Let P_s be a point of the path which is near enough to P_1 such that P_t for $s \leqq t \leqq 1$ are all contained in U. Also let W be a neighborhood of P_s, which is contained in U, such that ψ_s maps it into **D**. In this case $\psi_s \circ \phi^{-1}$ gives an analytic isometry from $\phi(W)$ to $\psi_s(W)$, which can be extended uniquely to a global isometry σ of **D** by Lemma 2. It is a simple matter to see that the germ ψ_1 determined by $\sigma \circ \phi$ at P_1 satisfies our demand.

Now we define a mapping Ψ from $\bar{V}$ into **D** by $\Psi(P_1) = \psi_1(P_1)$. Since $\psi_1(P_1)$ is independent on the homotopy of paths from P_0 to P_1 and since $\bar{V}$ is assumed to be simply connected, Ψ is single-valued completing the proof.

We note that the changes of P_0 and ψ_0 in the definition of Ψ amount to replace it by $\sigma \circ \psi$, where σ is an operation of the group **G**; hence the mapping Ψ is essentially unique.

Now we shall assume that $(\bar{V}^n, f)$ is the universal covering variety of a compact Kähler variety V^n with constant curvature. It is clear that the projection f induces a similar metric on $\bar{V}$. Therefore we can find a single-valued analytic isometry Ψ from $\bar{V}$ into the normal domain **D** according to Lemma 3. However in the present case the following holds.

THEOREM 1. *The mapping* Ψ *gives a homeomorphic analytic isometry of* $\bar{V}$ *onto* $\mathbf{D}$.[2]

Proof. We shall show that we can define a single-valued analytic isometry Ψ^* from $\mathbf{D}$ into $\bar{V}$ such that $\Psi \circ \Psi^* = \Psi^* \circ \Psi = 1$. Let ψ^*_0 be the inverse of ψ_0 in the proof of Lemma 3; let Q_1 be an arbitrary point of $\mathbf{D}$ and connect it to the origin of $\mathbf{D}$ by a differentiable path Q_t $(0 \leqq t \leqq 1)$. If ψ^*_t $(0 \leqq t < 1)$ is a family of germs of coherent isometries from $\mathbf{D}$ into $\bar{V}$ sending Q_t to P_t, we can complete it by ψ^*_1. In fact since the path $f(P_t)$ on V has a finite length and since V is compact, $\lim_{t\to 1} f(P_t)$ exists and is a point of V. Therefore $\lim_{t\to 1} P_t = P_1$ also exists and is a point of $\bar{V}$. It is clear that P_t $(0 \leqq t \leqq 1)$ is a continuous path on $\bar{V}$ and we have only to take the inverse of ψ_1 as ψ^*_1. Since $\mathbf{D}$ is simply connected, it follows from this argument that we can define a single-valued analytic isometry Ψ^* from $\mathbf{D}$ into $\bar{V}$, which "extends" ψ^*_0. The relation $\psi_0 \circ \psi^*_0 = \psi^*_0 \circ \psi_0 = 1$ is clearly preserved by this extension.

It is essential in this theorem that $\bar{V}$ is the universal covering variety of a *compact* variety.

3. Structures. A compact Kähler variety V^n with constant curvature will be called *elliptic*, *hyperbolic* or *parabolic* in an obvious manner. Our Theorem 1 asserts that the universal covering variety $\bar{V}^n$ is then S^n, K^n or C^n respectively. Moreover the Poincaré group Γ of V is represented as a subgroup of $\mathbf{G}$, which is "discontinuous" and has no fix-point in $\mathbf{D}$. Therefore

THEOREM 2. *The elliptic variety* V^n *is the complex projective space* S^n.

Proof. This follows at once from the fact that every projective transformation in S^n has $n+1$ fix-points.

The following lemma will simplify some discussions in the proof of Theorem 3.

LEMMA. *Let* M^n *be a compact differentiable manifold and let* f_p $(p = 1, 2, \cdots)$ *be a family of differentiable mappings from* M *into another differentiable manifolds* N_p *with the following two properties*: a) *The family of point sets* X_p *in* M *such that* f_p *is not "regular" at* x *in* X_p (the Jacobian

[2] The author was notified of Hawley's paper on the same subject ("Constant holomorphic curvature," *Canadian Journal of Mathematics*, vol. 5 (1953), pp. 53-56) by Professor Sasaki in Princeton.

matrix of f_p at x has rank $< n$) *is decreasing from a certain* p *onward and has an empty intersection.* b) *The family of point sets* Δ_p *in* $M \times M$ *such that* $f_p(x) = f_p(y)$ *is decreasing from a certain* p *onward and has* Δ *as its intersection. In this case* f_p *gives an analytic homeomorphism from* M *into* N_p *from a certain* p *onward.*

Proof. Since M is compact and since X_p are closed, a) implies that X_p is empty from a certain p onward. However $\Delta_p - \Delta$ is closed, when X_p is empty. Since $M \times M$ is compact, b) implies therefore that $\Delta_p = \Delta$ from a certain p onward.

THEOREM 3. *In the hyperbolic case we can define an analytic homeomorphism* F *from* V^n *onto a non-singular variety in a complex projective space by a set of automorphic forms of the same weight in* $\mathbf{D}$ *with respect to* Γ.

Proof. (cf. [9], chap. 10). Let $j_\sigma(z)$ be the Jacobian of the isometry σ of $\mathbf{D}$ at z, then the series $\sum_{\sigma \in \Gamma} |j_\sigma(z)|^m$ $(m \geqq 2)$ is uniformly convergent on every compact set in $\mathbf{D}$. Moreover since the factor space $\mathbf{D}/\Gamma = V$ is compact, the set $\mathbf{F}$ of points z in $\mathbf{D}$ such that $|j_\sigma(z)| \leqq 1$ for all σ in Γ is also compact and clearly generates $\mathbf{D}$. Let r_0 be the number of σ in Γ such that $\mathbf{F}^\sigma$ intersects with $\mathbf{F}$; take $r \geqq 2r_0 - 1$ and put $N = C_n^{n+r} - 1$, then the set of "Poincaré series" $\Phi_m(z, \phi) = \sum_{\sigma \in \Gamma} \phi(\sigma z) \cdot j_\sigma(z)^m$ $(m \geqq 2)$ defines an analytic mapping Φ_m from $\mathbf{D}/\Gamma$ into the projective space S^N when $\phi(z)$ varies in the space of polynomials of degree not greater than r. Thereby if z, z_1 and z_2 are given points of $\mathbf{F}$ such that z_1 and z_2 are inequivalent under Γ, then Φ_m is regular at z and $\Phi_m(z_1) \neq \Phi_m(z_2)$ from a certain m onward. Hence we can apply the previous lemma to $f_p = \Phi_2 \times \Phi_3 \times \cdots \times \Phi_{p+1}$ and we conclude that f_p gives an analytic homeomorphism from $\mathbf{D}/\Gamma$ into p-ply projective space $S^N \times S^N \times \cdots \times S^N$ from a certain p onward. Therefore if we denote by F the composition of one of such mappings and the well-known imbedding of multiply projective space into an ordinary projective space, the image $F(V)$ of V is an everywhere algebroid variety without multiple point.

Now we can apply a result of Chow ([5], Theorem 5) to $F(V)$, which is thus an algebraic variety. This idea is general and could be applied to a simplification of the theory of automorphic functions.

We shall now consider the parabolic case. The following theorem shows that a complex torus is "universal" in a certain sense.

THEOREM 4. *When* V^n *is of parabolic type, there exists a uniquely determined finite "normal" covering variety* T^n *of* V^n *which is a complex*

torus. Moreover this extension has only a finite number of possibilities depending on n.

Proof. As we have remarked before V^n is a factor space of C^n by a discontinuous group Γ of "motions" in C^n. Thereby since V is compact, Γ is "indecomposable." Therefore by a result of Frobenius ([6], p. 663) translations in Γ form a normal subgroup $\mathbf{T}$ of finite index with $2n$ independent generators. Moreover the extension Γ of $\mathbf{T}$ does not depend on the choice of the coordinates in C^n and is divided into a finite number of classes depending on n. We have now to put $T^n = C^n/\mathbf{T}$.

It follows from the above proof that the totality of parabolic varieties V^n forms a finite number of "analytic families" in a natural way such that two members in the same family have the same real analytic structure. In particular those which are algebraic form a finite number of analytic subfamilies. This is a consequence of the "theorem of Riemann-Weierstrass" and of the following general

THEOREM 5. *Let V^n and $\tilde{V}^n$ be complex varieties such that one of them is a finite covering variety of the other, then they are or they are not analytically homeomorphic to non-singular projective models at the same time.*

Proof. It follows from a general theorem of Kodaira ([8], Theorem 3) that a finite covering variety of a non-singular projective model is a non-singular projective model. Therefore we have only to prove the converse. Let $\tilde{V}$ be a non-singular subvariety of a projective space S^N, which is at the same time a finite covering variety of V with projection f. Let P be a point of V and let $F(P)$ be the "Chow point" of the 0-cycle $f^{-1}(P)$ in S^N which is carried by $\tilde{V}$, then we see readily that F is homeomorphic, analytic and regular.

4. Minimality. In this section we shall show that hyperbolic and parabolic varieties are "minimal" in a certain sense. We first put a general definition. (This definition is given essentially by Weil in his book on Abelian variety [Hermann, 1948] on page 27.)

DEFINITION. *Let V^n be a projective model and F a meromorphic mapping from a complex variety U^m into V^n, which can be defined invariantly using local coordinates on U and affine representatives of V. If every such F is necessarily analytic for any U, we say that V is a minimal model.*

We shall now prove the following remarkable

THEOREM 6. *Any hyperbolic variety is a minimal model. If a parabolic variety is a non-singular projective model, it is also a minimal model.*

Proof. Let F be a meromorphic mapping from U^m into a variety V^n which is stated in the theorem, then F is analytic. In order to prove this assertion we may assume that U^m is a connected neighborhood of the origin of C^m. Since V^n is imbedded in a projective space, the "fundamental locus" W of F have dimension at most equal to $m-2$; hence its residual space $U-W$ is simply connected. Let z be a point of $U-W$ and let P^* be a point of the universal covering variety $(\mathbf{D}, f)$ of V^n lying over $F(z)$, then the germ of analytic mapping $z \to P^*$ can be extended as a single-valued analytic mapping F^* from $U-W$ into $\mathbf{D}$. Thereby the simply connectedness of $U-W$ is essential. Now any component F^*_α of F^* is a bounded holomorphic function on $U-W$ for $\alpha = 1, 2, \cdots, n$. Therefore by a well-known theorem of Hartogs (cf. [3], Theorem 5, pp. 173-174) F^*_α must be holomorphic throughout U for $\alpha = 1, 2, \cdots, n$.

As a consequence of this theorem we shall prove a complete generalization of the theorem of Schwarz-Klein.

THEOREM 7. *Let* $\mathbf{K}$ *be the field of meromorphic* (hence algebraic) *functions on a hyperbolic variety* V^n; *then the group of automorphisms of* $\mathbf{K}$ *over* C *is finite.*

Proof. We know that every such an automorphism can be represented as a birational correspondence between V and itself. However since V is a minimal model, the birational correspondence is necessarily everywhere biregular; hence it is induced by some analytic homeomorphism of $\mathbf{D}$. However by a remark at the end of section 1 such operations form a subgroup of $\mathbf{G}$, which is discrete by a result of Bochner ([1], Theorem 5). On the other hand every discrete subgroup of $\mathbf{G}$ is discontinuous on $\mathbf{D}$ (cf. [9], 49); hence our assertion.

5. Further results. Let g_p $(1 \leqq p \leqq n)$ be the p-th *irregularity* of a compact complex variety V^n, which is the number of linearly independent p-ple differentials of the first kind; we put also $g_0 = 1$. In analogy with the Euler-Poincaré characteristic $\chi(V^n)$, we call $\chi_a(V^n) = \sum_{p=0}^{n} (-1)^p g_p$ the *analytic characteristic* of V^n.

Now we consider a hyperbolic variety V^n. If we denote by dv the

invariant measure in its normal domain **D** which coincides with the Euclidean "volume element" at the origin of **D**, the total measure $v(V^n)$ of V^n has an intrinsic meaning and we have the following

THEOREM 8. *It holds that*

$$\chi(V^n) = (n+1)\cdot\chi_a(V^n) = (-1)^n((n+1)!/\pi^n)v(V^n).$$

Proof. If we remember Chern's generalization of Gauss-Bonnet's formula ([4], pp. 114-115) which is valid in general and the so-called "Todd's formula" which was proved recently by Hirzebruch for non-singular projective models, we have only to verify this identity (except for $(-1)^n$) in the case of the projective space S^n.

We next consider a parabolic variety V^n and we shall prove the following

THEOREM 9. *There exists a uniquely determined series of compact complex varieties $V = V_0, V_1, \cdots, V_{s+1} = W$ such that V_i is a fibre space over its "Albanese variety" T_i* (complex torus attached to the period matrix of simple differentials of the first kind on V_i) *with fibre V_{i+1} for $i = 0, 1, \cdots, s$ The final variety W is a parabolic variety with the first irregularity equal to zero.*

Proof. We can write Γ in the form $z^* = Az + a$ and consider A as a representation of Γ in $U(n)$ with kernel $\mathbf{T}$. The factor group $G = \Gamma/\mathbf{T}$ is finite and gives the group of "covering transformations" of $T = C^n/\mathbf{T}$ over V. It is clear that the projection from T onto V induces a natural isomorphism between the module of differentials of the first kind on V and the module of invariant differentials with constant coefficients in dz^α $(1 \leqq \alpha \leqq n)$ under the operation of G. In particular g_1 is nothing but the number of principal characters in A. Therefore by a suitable coordinate transformation in C^n, the operation of Γ can be represented in the form

$$\begin{Bmatrix} I_{g_1} & 0 & a_0 \\ 0 & A_1 & a_1 \\ 0 & 0 & 1 \end{Bmatrix} \quad (I_g = \text{unit matrix})$$

and it is easy to see that the vectors a_0 form a discrete group of rank $2g_1$ in C^{g_1} which determines a torus T_0. If we consider the partial matrix $\begin{Bmatrix} A_1 & a_1 \\ 0 & 1 \end{Bmatrix}$ corresponding to $a_0 = 0$, we can repeat the same reasoning.

The following corollary shows that the "arithmetic genus" of a parabolic variety V^n is equal to $(-1)^{n-1}$.

COROLLARY. *It holds that* $\chi(V^n) = \chi_a(V^n) = 0$.

In fact we conclude from the above proof that

$$\sum_{p=0}^{n} g_p t^{n-p} = (1/(G)) \sum_{A \in G} \det(tI_n + A)$$

with an indeterminate t, where (G) is the order of G. However since every transformation in Γ has no fix-point, we have $\det(A - I_n) = 0$; whence $\sum g_p(-1)^{n-p} = 0$.

On the other hand it would be favorable, if a variety W stated in Theorem 9 does not exist actually. However the circumstance is not so simple.

EXAMPLE. Let M be a discrete additive group of rank 2 in C; let x_α, y_α $(1 \leqq \alpha \leqq 3)$ be six complex numbers satisfying the following conditions: $x_1, y_2, x_3 - y_3 \not\equiv 0 \pmod{M}$, $2x_1, 2y_2, 2(x_3 - y_3) \equiv 0 \pmod{M}$; put $\sigma_1 : z^*_1 = z_1 + x_1$, $z^*_2 = -z_2 + x_2$, $z^*_3 = -z_3 + x_3$ and $\sigma_2 : z^*_1 = -z_1 + y_1$, $z^*_2 = z_2 + y_2$, $z^*_3 = -z_3 + y_3$. Let $\mathbf{T}$ be a group of translations of the form $\tau : z^*_\alpha = z_\alpha + m_\alpha$ $(m_\alpha \in M)$ $(1 \leqq \alpha \leqq 3)$, which generates a discontinuous group Γ of motions in C^3 with σ_1 and σ_2. It can be readily verified that every member of Γ has no fix-point in C^3 and that the factor group $\Gamma/\mathbf{T}$ is the Klein's four group. Since the factor space $T^3 = C^3/\mathbf{T}$ is a product of three elliptic curves corresponding to M, the factor space $V^3 = C^3/\Gamma$ is also a non-singular projective model by Theorem 5 which is covered four times by T^3. Moreover the Poincaré polynomial of V^3 is given by $1 + 3t^2 + 8t^3 + 3t^4 + t^6$.

However in the case of curves and surfaces, we have $g_1 \geqq 1$ by the above corollary. The case of $n = 1$ is trivial and gives an elliptic curve. The case of $n = 2$ is either a complex torus or is a fibre space over an elliptic curve with another elliptic curve as its fibre by Theorem 9. In the latter case they are necessarily non-singular projective models and known as "hyperelliptic surfaces."

HARVARD UNIVERSITY AND KYOTO UNIVERSITY, JAPAN.

BIBLIOGRAPHY.

[1] S. Bochner, "Vector fields and Ricci curvature," *Bulletin of the American Mathematical Society*, vol. 52 (1946).

[2] ———, "Curvature in Hermitian metric," *Bulletin of the American Mathematical Society*, vol. 53 (1947).

[3] ——— and W. T. Martin, *Several complex variables*, Princeton University Press (1948).

[4] S. S. Chern, "Characteristic classes of Hermitian manifolds," *Annals of Mathematics*, vol. 47 (1946).

[5] W. L. Chow, "On compact complex analytic varieties," *American Journal of Mathematics*, vol. 71 (1949).

[6] G. Frobenius, "Ueber die unzerlegbare diskreten Bewegungsgruppen," Preussische Akademie der Wissenschaften Sitzungsberichte, vol. 1 (1911).

[7] E. Kähler, "Ueber eine bemerkenswerte Hermitische Metrik," *Hamburg Seminar für Mathematik Abhandlungen*, vol. 9 (1933).

[8] K. Kodaira, "On Kähler varieties of restricted type," to appear in *Proceedings of the National Academy of Sciences*.

[9] C. L. Siegel, *Analytic functions of several complex variables*, Princeton Lecture Notes (1950).

LOWER SEMI-CONTINUITY AND AREA FUNCTIONALS, II. THE BANACH AREA.* [1]

By CASPER GOFFMAN.

1. Introduction. In part I of this series [1], we showed that the Lebesgue area for continuous non-parametric surfaces has a natural extension to a lower semi-continuous functional on the space of summable functions suitably metrized. Specifically, if

$$d(p, q) = \int_0^1 \int_0^1 | p(x, y) - q(x, y) | \, dx \, dy$$

is taken as the distance between polyhedral functions $p(x, y)$ and $q(x, y)$, the extension of elementary area to the space of equivalence classes of summable functions, obtained by the Fréchet process [2], is a lower semi-continuous functional. It agrees with the Lebesgue area for continuous surfaces and, in general, retains the character of the Lebesgue area.

The present paper is an attempt to obtain similar results for the parametric case. We are concerned with the following existence problem:

Is there a distance function $d(p, q)$ for the parametric polyhedral mappings such that

(a) elementary area is lower semi-continuous on the space $\mathfrak{P}$ defined by $d(p, q)$,

(b) the completion $\mathfrak{P}'$ of $\mathfrak{P}$ contains a reasonably wide class of discontinuous mappings along with all the continuous ones?

We find that the analogue of the metric used in the non-parametric case is unsatisfactory in the parametric case since the elementary area $E(p)$ is not lower semi-continuous. Nevertheless, we find a distance function for which $\mathfrak{P}'$ contains all Baire mappings (i. e. mappings $f: x_i(u, v)$, $i = 1, 2, 3$ for which $x_i(u, v)$ are Baire functions) whose images on the coordinate planes have finite planar measure. We remark that the extended functional obtained here is related to the Banach definition of area rather than to the Lebesgue definition.

* Received January 21, 1954.

[1] Research supported by a grant from the National Science Foundation.

2. Natural metrics for which $E(p)$ is not lower semi-continuous. The analogue of the above metric for parametric polyhedral mappings $p: p_i(u, v)$ and $q: q_i(u, v)$, $i = 1, 2, 3$ is the function

$$d(p, q) = \int_0^1 \int_1^0 \sum_{i=1}^{3} | p_i(u, v) - q_i(u, v) | \, du \, dv.$$

This metric cannot be used as a basis for a theory of surface area since the elementary area $E(p)$ is not lower semi-continuous on the resulting space.[2] Indeed, consider the mapping $p: p_1(u, v) = u$, $p_2(u, v) = v$, $p_3(u, v) = 0$ on the unit square I. Subdivide I, for every positive integer n, into a grill of squares $I_{n1}, I_{n2}, \cdots, I_{nk_n}$, all of whose diagonals are of length $1/2^n$. Place in each of these squares, I_{nk}, a pair of squares J_{nk} and K_{nk}, concentric with I_{nk}, such that the sum of the areas of the J_{nk} is $1\text{-}1/2^n$ and the sum of the areas of the K_{nk} is $1/2^n$. Now, let K_n be a simply connected polygonal region in I which contains all the K_{nk} in its interior and whose area is less than $1/n$. There is a homeomorphism $f_n: f_{ni}(u, v)$, $i = 1, 2, 3$, between I and K_n which takes each J_{nk} into K_{nk}. There is then a polyhedral mapping $q: q_{ni}(u, v)$, $i = 1, 2, 3$, such that $| p_i(u, v) - q_{ni}(u, v) | < 1/2^n$ for each i on a set of measure greater than $1 - 1/2^n$, and such that $E(q_n) < 1/2^n$. Now,

$$\lim_{n \to \infty} \int_0^1 \int_0^1 \sum_{i=1}^{3} | p(u, v) - q_{ni}(u, v) | \, du \, dv = 0.$$

But, $\lim_{n \to \infty} E(q_n) = 0$ and $E(p) = 1$, so that the elementary area is not lower semi-continuous.

It was also noted in [1] that if the distance between continuous non-parametric surfaces is taken as

$$\delta(f, g) = m(E),$$

where E is the set for which $f(x, y) \neq g(x, y)$, then the Lebesgue area is lower semi-continuous and the Fréchet process leads to a lower semi-continuous extension to the set of all measurable functions. We now show that the elementary area also fails to be lower semi-continuous for the parametric analogue of this metric. Indeed, let $\phi(x)$ be the Cantor function on $[0, 1]$ and let f be given by

$$x_1(u, v) = u, \qquad x_2(u, v) = v, \qquad x_3(u, v) = \phi(u).$$

It is clear that the Lebesgue area $L(f)$ of f exceeds 1. Moreover, it is easily

[2] E. Silverman has shown recently [3] that the Lebesgue area is lower semi-continuous with respect to point wise convergence.

seen that, for every $\epsilon > 0$, there is a continuous f_ϵ such that $\delta(f, f_\epsilon) < \epsilon$ and $L(f_\epsilon) < 1$. This shows that the functional $L(f)$ is not lower semi-continuous on the given space. We remark, however, that the elementary area is lower semi-continuous on the set of parametric mappings of class C' with $\delta(f, g)$ as metric. This leads to an area functional defined on a wide class of mappings which does not agree with the Lebesgue area for certain continuous mappings. We shall discuss this case elsewhere.

3. The space $\mathfrak{P}$. We now introduce a distance function for the polyhedral mappings for which the elementary area $E(p)$ is lower semi-continuous. For convenience, we shall work first with transformations $T: x(u, v), y(u, v)$ of the unit square $I = [0, 1] \times [0, 1]$ into the (x, y)-plane and shall speak of mappings (triples of transformations) later. Our distance function is defined in terms of a special sequence of subdivisions of I. For every n, let

$$R_1^n, R_2^n, \cdots, R_{2^{2n}}^n$$

be the set of squares in I, whose sides are of length $1/2^n$, obtained by means of the lines $u, v = 0, 1/2^n, \cdots, 1 - 1/2^n, 1$, where each bounding edge is assigned to one of the two squares which it bounds in such a way that each R_j^{n+1} is contained in an R_k^n.

Now, let P and Q be polyhedral transformations of I into the (x, y)-plane; i. e., I may be decomposed into a finite set of triangles such that P is continuous on I and is linear on each triangle of the decomposition, and similarly for Q. Using TR as notation for the image set of a set R under a transformation T, and $R * S$ as notation for the symmetric difference $R \cup S - R \cap S$ of sets R and S, we write $\mu_i^n(P, Q) = m(PR_i^n * QR_i^n)$ for the planar measure of $PR_i^n * QR_i^n$.

We shall say that a sequence $\{P_n\}$ of polyhedral transformations converges to a polyhedral transformation P if for every $i = 1, 2, \cdots, 2^{2m}$; $m = 1, 2, \cdots$, we have $\lim_{n \to \infty} \mu_i^m(P, P_n) = 0$. The neighborhoods U of P which correspond to this notion of convergence are given by finite sequences $(m_1, i_1), (m_2, i_2), \cdots, (m_k, i_k)$; $\epsilon_1, \epsilon_2, \cdots, \epsilon_k$, where the ϵ's are positive, and $Q \in U$ if $\mu_{i_j}^{m_j}(P, Q) < \epsilon_j$ for every $j = 1, 2, \cdots, k$. It then follows that the space $\mathfrak{P}$ obtained in this way is such that

THEOREM 1. *$\mathfrak{P}$ is a uniform space which satisfies the Hausdorff first denumerability condition and so it is metrizable.*

It is convenient to work with an equivalent distance function $d(P, Q)$ instead of the given convergence notion. For every n, we let

$$d_n(P, Q) = \sum_{i=1}^{N} \mu_i^n(P, Q) \text{ where } N = 2^{2n}.$$

The function $d_n(P, Q)$ is monotonically non-decreasing in n. We let $d(P, Q) = 1/2^n$ where n is the smallest integer for which $d_n(P, Q) \geqq 1/2^n$.

LEMMA 1. *A sequence $\{P_n\}$ converges to P in $\mathfrak{P}$ if and only if*

$$\lim_{n\to\infty} d(P, P_n) = 0.$$

Proof. Suppose $\{P_n\}$ converges to P in $\mathfrak{P}$. For every n, there is an N such that, for every $m > N$, and $i = 1, 2, \cdots, 2^{2n}$, we have $\mu_i^n(P, P_m) < (\frac{1}{2})^{3n}$, so that $d_n(P, P_m) < 1/2^n$, whereby $d(P, P_m) < 1/2^n$.

Conversely, suppose $\{P_n\}$ does not converge to P in $\mathfrak{P}$. There is then an i and m with $\limsup_{n\to\infty} \mu_i^m(P, P_n) > 1/2^k > 0$. Moreover, we may choose $k > m$. Then, for every N, there is an $n > N$ with $d_m(P, P_n) \geqq \mu_i^m(P, P_n) > 1/2^k$. Since $k > m$, we have $d_k(P, P_n) \geqq d_m(P, P_n) > 1/2^k$, so that $d(P, P_n)\ 1/2^k$.

We mention, in passing, that $d(P, Q)$ is not a metric but satisfies (i) $d(P, Q) > 0$ if $P \neq Q$, (ii) $d(P, P) = 0$, (iii) $d(P, Q) = d(Q, P)$, and (iv) $d(P, Q) + d(Q, R) \geqq \frac{1}{2} d(P, R)$.

The elementary area $E(p)$ of a polyhedral transformation P has its usual meaning [4].

THEOREM 2. *The functional $E(p)$ is lower semi-continuous on the space $\mathfrak{P}$.*

Proof. Observe that for every polyhedral P and positive integer n we have $E(P) \geqq \sum_{i=1}^{N} m(PR_i^n)$ where $N = 2^{2n}$, but that for every $\epsilon > 0$ there is an $n = n(P, \epsilon)$, with $1/2^n < \epsilon/2$, for which $E(P) < \sum_{i=1}^{N} m(PR_i^n) + \epsilon/2$. Suppose $d(P, Q) < 1/2^n$ where $n = n(P, \epsilon)$. Then $d_n(P, Q) < 1/2^n$, so that

$$\begin{aligned} E(Q) &\geqq \sum_{i=1}^{N} m(QR_i^n) \geqq \sum_{i=1}^{N} m(PR_i^n) - \sum_{i=1}^{N} \mu_i^n(P, Q) \\ &= \sum_{i=1}^{N} m(PR_i^n) - d(P, Q) > \sum_{i=1}^{N} m(PR_i^n) - \epsilon/2 > E(P) - \epsilon. \end{aligned}$$

This proves the lower semi-continuity.

It is also clear that for every P there is a sequence $\{P_n\}$, $P_n \neq P$, such that $\lim_{n\to\infty} d(P, P_n) = 0$ and $E(P) = \lim_{n\to\infty} E(P_n)$. Thus, by the Fréchet pro-

cess, $E(P)$ has a lower semi-continuous extension to the completion $\mathfrak{P}'$ of the space $\mathfrak{P}$.

We also wish to remark that $E(P)$ is not continuous on $\mathfrak{P}$; examples of this fact are easy to construct and we shall omit them. In this connection we make the conjecture that there is no metric for the polyhedral mappings for which $E(P)$ is continuous for which the completed space contains all continuous mappings.

4. Measurable point and set transformations. The completion $\mathfrak{P}'$ of $\mathfrak{P}$ involves the following sort of set functions with the rectangles R_i^n, $i = 1, 2, \cdots, 2^{2n}$; $n = 1, 2, \cdots$ as domain. Associated with every R_i^n there is a measurable set $S_i^n = f(R_i^n)$ in the (x, y)-plane such that $S_i^n = \bigcup_{j=1}^{4} S_{ij}^{n+1}$, where the R_{ij}^{n+1}, $j = 1, \cdots, 4$, are the four squares whose union is R_i^n. The set function $f(R_i^n)$ has an associated point function $\phi(x, y)$ defined as follows: For each n, let $\phi_n(x, y)$ be the number of squares R_i^n, $i = 1, 2, \cdots, 2^{2n}$, for which $(x, y) \in f(R_i^n)$, and let $\phi(x, y) = \lim_{n \to \infty} \phi_n(x, y)$ or ∞, as the case may be. The function $\phi(x, y)$ is measurable. For, the set of points for which $\phi_n(x, y) = k$, for every n and $0 \leqq k \leqq n$, is expressible as a finite number of unions and intersections of the sets $f(R_i^n)$, $i = 1, 2, \cdots, 2^{2n}$, and so it is measurable. In the context, when we refer to set functions we shall always mean those of the above type.

By the Banach area of a set function f we mean the number

$$B(f) = \int\int \phi(x, y)\, dx\, dy,$$

where the integration is over the whole (x, y)-plane. We shall show, in the next section, that the completion $\mathfrak{P}'$ of $\mathfrak{P}$ is just the set of all set functions f for which the associated $\phi(x, y)$ is different from 0 on a set of finite measure.

We shall call a point transformation T measurable if for every Borel set S, the image TS is measurable. Every measurable point transformation has an associated set transformation f defined by $f(R_i^n) = TR_i^n$ which evidently satisfies the above condition. The Banach area $B(T)$ is defined as $B(f)$ where f is the set function which corresponds to T. On the other hand, there are set functions $S_i^n = f(R_i^n)$ which cannot be obtained from point transformations [5] in the above way.

For every n, let $B_n(f) = \sum_{i=1}^{N} m(S_i^n)$. We have the following fact, essentially proved by Banach [6].

THEOREM 3. $B(f) = \lim_{n\to\infty} B_n(f)$.

Proof. $B_n(f) = \int\int \phi_n(x,y)\,dx\,dy$ and $\{\phi_n(x,y)\}$ is a monotonically non-decreasing sequence converging to $\phi(x,y)$.

5. The space $\mathfrak{P}'$. We show now that the completion $\mathfrak{P}'$ of $\mathfrak{P}$ consists of the set $\mathfrak{F}$ of all set functions f for which the set $f(I)$ has finite measure, and that the lower semi-continuous extension of $E(P)$ to $\mathfrak{P}'$ agrees with $B(f)$ on $\mathfrak{F}$.

We first note that the definition of distance $d(f,g)$ given in section 3 for polyhedral transformations can be carried over without change to the set $\mathfrak{F}$; we shall henceforth designate this space by $\mathfrak{F}$.

(a) *$\mathfrak{F}$ is complete.* For, let $\{f_n\}$ be a fundamental sequence in $\mathfrak{F}$. For each R_i^n, consider the sequence $\{f_m(R_i^n)\}$ of measurable sets. This sequence is fundamental in the sense that for every $\epsilon > 0$ there is an N such that $m,k > N$ implies $m(f_k(R_i^n) * f_m(R_i^n)) < \epsilon$. There is a subsequence $\{f_{i_j}\}$ such that $m(f_{i_{j+1}}(R_i^n) * f_{i_j}(R_i^n) < 1/2^j$, for every j. Let $S_i^n = \bigcup_{k=1}^{\infty} \bigcap_{j=k}^{\infty} f_{i_j}(R_i^n)$ and define the functions $S_i^n = f(R_i^n)$. It is a routine matter to check that f is a set function and that $\lim_{n\to\infty} d(f,f_n) = 0$.

(b) *$\mathfrak{P}$ is dense in $\mathfrak{F}$.* Let $f \in \mathfrak{F}$ be given by $S_i^n = f(R_i^n)$. For any n, there are polygons Q_i^n, $i = 1, \cdots, 2^{2n}$, such that $m(Q_i^n * S_i^n) < 2^{-(3n+1)}$. There is a polyhedral P such that $P(R_i^n) * Q_i^n < 2^{-(3n+1)}$ for every i, so that

$$d_n(P,Q) = \sum_{i=1}^{N} m(P(R_i^n) * f(R_i^n)) \leqq \sum_{i=1}^{N} m(P(R_i^n) * Q_i^n) + \sum_{i=1}^{N} m(Q_i^n * S_i^n) < 1/2^n,$$

where $N = 2^{2n}$ and $d(P,f) < 1/2^n$, so that $\mathfrak{P}$ is dense in $\mathfrak{F}$. Steps (a) and (b) furnish the proof of

THEOREM 4. *$\mathfrak{F}$ is the completion of $\mathfrak{P}$, so that $\mathfrak{F} = \mathfrak{P}'$.*

The proof that $B(f)$ is the Fréchet extension of $E(P)$ to $\mathfrak{P}'$ will also be given in two steps.

(a) *$B(f)$ is lower semi-continuous on $\mathfrak{P}'$.* Suppose $B(f) < \infty$. There is an n such that $B_n(f) > B(f) - (\frac{1}{2})^{n+1}$. Now, if $d(g,f) < (\frac{1}{2})^{n+1}$, then $d_n(g,f) < (\frac{1}{2})^{n+1}$, so that $B(g) \geqq B_n(g) > B_n(f) - (\frac{1}{2})^{n+1} > B(f) - 1/2^n$.

If $B(f) = \infty$, then for every M there is an n such that $B_n(f) > M + 1$, so that if $d(g, f) < (\frac{1}{2})^{n+1}$ then again $B(g) \geqq B_n(g) > B_n(f) - 1 > M$.

(b) *For every $f \in \mathfrak{P}'$ there is a sequence $\{P_n\}$ of polyhedral transformations such that* $\lim_{n\to\infty} d(f, P_n) = 0$ *and* $\lim_{n\to\infty} E(P_n) = B(f)$. To prove this we simply observe that the polyhedral P in (b) above may be taken so that $E(P) < B_n(f) - 1/2^n$. Writing P_n for this P we have $\lim_{n\to\infty} d(P_n, f) = 0$ and $\lim_{n\to\infty} E(P_n) \leqq \lim_{n\to\infty} B_n(f) = B(f)$, by Theorem 3. But $\lim_{n\to\infty} E(P_n) \geqq B(f)$ by the lower semi-continuity of $B(f)$.

Now, the content of the Fréchet process is that there is a unique extension of a lower semi-continuous functional on a metric space to its completion with properties (a) and (b). We have thus proved

THEOREM 5. *The unique extension to $\mathfrak{P}'$ of $E(P)$ on $\mathfrak{P}$ given by the Fréchet process is the Banach area $B(f)$.*

We conclude this section with a remark on an analogue of the Fréchet equivalence relation [7] for the case under consideration. The discussion will be restricted to the subspace $\mathfrak{T} \subset \mathfrak{P}'$ of point transformations since if f is a set function and τ is a one-one transformation between I and itself, even polyhedral, then $f\tau$ may be meaningless. We shall say that T_1 is equivalent to T_2 if, for every $\epsilon > 0$ there are one-one transformations τ_1 and τ_2 between I and itself such that $d(T_1, T_2\tau_2) < \epsilon$ and $d(T_1\tau_1, T_2) < \epsilon$. It is evident that this is an equivalence relation. Writing $\phi(x, y; T)$ for the multiplicity function of the transformation T, we prove that T_1 is equivalent to T_2 if and only if $\phi(x, y; T_1) = \phi(x, y; T_1)$ almost everywhere; in other words, surfaces may be identified with multiplicity functions.

LEMMA 2. *If* $\lim_{n\to\infty} d(T, T_n) = 0$ *then* $\liminf_{n\to\infty} \phi(x, y; T_n) \geqq \phi(x, y; T)$ *almost everywhere.*

Proof. We need only observe that for every m, $\liminf_{n\to\infty} \phi(x, y; T_n) \geqq \phi_m(x, y; T)$ almost everywhere. This holds since for every $\epsilon > 0$ there is an N such that $n > N$ implies $m(T_n(R_i^m) * T(R_i^m)) < \epsilon/2^{2m}$ for every $i = 1, 2, \cdots, 2^{2m}$, and a standard real variable argument supplies the proof.

COROLLARY 1. *If T_1 is equivalent to T_2 then $\phi(x, f; T_1) = \phi(x, y; T_2)$ almost everywhere.*

Proof. For every τ, it is evident that $\phi(x, y; T_1\tau) = \phi(x, y; T_1)$. Let $\{\tau_n\}$

be such that $\lim_{n\to\infty} d(T_1\tau_n, T_2) = 0$. Then $\liminf_{n\to\infty} \phi(x, y; T_1\tau_n) \geqq \phi(x, y; T_2)$ so that $\phi(x, y; T_1) \geqq \phi(x, y; T_2)$, almost everywhere. Similarly $\phi(x, y; T_2) \geqq \phi(x, y; T_1)$ almost everywhere.

The converse also holds.

LEMMA 3. *If $\phi(x, y; T_1) = \phi(x, y; T_2)$ almost everywhere then T_1 is equivalent to T_2.*

Proof. We need only show that for every $\epsilon > 0$ there is a τ such that $d(T_1\tau, T_2) < \epsilon$. Let $1/2^n < \epsilon$, and consider the sets $T_2(R_i^n)$, $i = 1, 2, \cdots, 2^{2n}$. Since $\phi(x, y; T_1) = \phi(x, y; T_2)$, there are sets S_i, $i = 1, 2, \cdots, 2^{2n}$, such that $T_1(S_i) = T_2(R_i^n)$ for every i. Now, let τ be a one-one transformation which takes each R_i^n onto S_i. Then $T_1\tau(R_i^n) = T_1(S_i) = T_2(R_i^n)$ for every i so that $d(T_1\tau, T_2) < \epsilon$.

Remark. In the above proof, if S_i should be of cardinality less than c the necessary adjustment may be made by adjoining to it a set of cardinality c whose image under T_1 has measure 0.

THEOREM 6. *Two transformations T_1 and T_2 are equivalent if and only if they have the same multiplicity function.*

6. Mappings in 3-space. Most of the above results on mappings in 2-space go over with little change to mappings in 3-space. Thus, the completion $\mathfrak{P}'$ of the space $\mathfrak{P}$ of polyhedral mappings $p: p_i(u, v)$, $i = 1, 2, 3$, contains all mappings $x: x_i(u, v)$, $i = 1, 2, 3$, for which the $x_i(u, v)$ are Baire functions, and the Banach area $B(x)$ is lower semi-continuous on $\mathfrak{P}'$. However, we do not know whether $B(f)$ always agrees with the Fréchet extension $\Phi(x)$ of $E(p)$ to $\mathfrak{P}'$. We know only that $B(x) \leqq \Phi(x)$ always, and that $B(x) = \Phi(x)$ if $\Phi(x)$ is infinite. We proceed with the details.

A polyhedral mapping $p: p_i(u, v)$, $i = 1, 2, 3$, into (x_1, x_2, x_3)-space has three associated polyhedral transformations $P_i(u, v)$, $i = 1, 2, 3$, some of which may be degenerate, where $P_i(u, v)$ is the pair $(p_{i+1}(u, v), p_{i+2}(u, v))$ and all integers are taken module 3. Let $\mathfrak{P}$ be the space of all polyhedral mappings with $d(p, q) = \sum_{i=1}^{3} d(P_i, Q_i)$; clearly, $\mathfrak{P}$ is metrizable.

A set mapping $f: f_i$, $i = 1, 2, 3$, is a triple of set functions. We show that the completion $\mathfrak{P}'$ of $\mathfrak{P}$ is the space $\mathfrak{F}$ of all set mappings f for which the images of the f_i have finite measure, and where $d(f, g) = \sum_{i=1}^{3} d(f_i, g_i)$. The proof that $\mathfrak{F}$ is complete is the same as for set functions. We need

accordingly only show that $\mathfrak{P}$ is dense in $\mathfrak{F}$. For this, we fix n, and for each R_j^n, $j = 1, 2, \cdots, 2^{2n}$, and $i = 1, 2, 3$, consider the set $f_i(R_j^n)$ in the (x_{i+1}, x_{i+2})-plane. Now, this is a measurable set of finite measure so that there is a polygon Q_{ij} in the (x_{i+1}, x_{i+2})-plane for which $m(f_i(R_j^n) * Q_{ij}) < (\frac{1}{2})^{2n+2}$. We construct a polyhedral mapping p such that the image of each R_j^n contains the three polygons Q_{ij}, $i = 1, 2, 3$, and $E(p) < \sum_{i=1}^{3} \sum_{j=1}^{N} m(Q_{ij}) < (\frac{1}{2})^{n+2}$, where $N = 2^{2n}$. An easy calculation shows that $d(f, p) < 1/2^n$, so that $\mathfrak{F} = \mathfrak{P}'$.

For every $f \in \mathfrak{F}$, we let

$$B_n(f) = \sum_{i=1}^{N} \{[m(f_1(R_i^n)]^2 + [m(f_2(R_i^n)]^2 + [m(f_3(R_i^n)]^2\}^{\frac{1}{2}}.$$

and define the Banach area by $B(f) = \lim_{n\to\infty} B_n(f)$. We show that $B(f)$ is lower semi-continuous on $\mathfrak{F} = \mathfrak{P}'$. Suppose $f \in \mathfrak{P}'$ and $B(f) < \infty$. Let n be such that $B_n(f) > B(f) - \epsilon/2$ and $1/2^n < \epsilon/6$. If $d(f, g) < 1/2^n$ then $\sum_{i=1}^{3} d(f_i, g_i) < 1/2^n$ so that $d_n(f_i, g_i) < 1/2^n$ for every $i = 1, 2, 3$. In other words, $\sum_{i=1}^{3} \sum_{j=1}^{N} m(f_i(R_j^n) * g_i(R_j^n) < 3/2^n$. But now,

$$|B_n(f) - B_n(g)| = |\sum_{j=1}^{N} \{\sum_{i=1}^{2} [m(f_i(R_j^n)]^2\}^{\frac{1}{2}} = \sum_{j=1}^{N} \{\sum_{i=1}^{3} [m(g_i(R_j^n))]^2\}^{\frac{1}{2}}|$$

$$\leqq \sum_{j=1}^{N} |\{\sum_{i=1}^{3} [m(f_i(R_j^n))]^2\}^{\frac{1}{2}} - \{\sum_{i=1}^{3} [m(g_i(R_j^n))]^2\}^{\frac{1}{2}}|$$

$$\leqq \sum_{j=1}^{N} \sum_{i=1}^{3} m(f_i(R_j^n) * g_i(R_j^n))$$

by the use of a well known inequality [8]. Thus,

$$B(g) \geqq B_n(g) \geqq B_n(f) - 3/2^n > B(f) - \epsilon/2 - 3\epsilon/6 = B(f) - \epsilon.$$

The proof of lower semi-continuity when $B(f) = \infty$ is similar and will be omitted. We observe that for polyhedral p, $B(p) = E(p)$, so that $B(f)$ is a lower semi-continuous extension of $E(p)$ to $\mathfrak{P}'$. Since the Fréchet extension $\Phi(f)$ is the maximal lower semi-continuous extension of $E(p)$ to $\mathfrak{P}'$, we have $\Phi(f) \geqq B(f)$.

We have already shown that if f is planar, then $\Phi(f) = B(f)$. We remark that if $\Phi(f)$ is infinite, then so is $B(f)$. This follows immediately from the obvious inequalities

$$B(f_i) \leqq B(f) \leqq \sum_{i=1}^{3} B(f_i) \text{ and } \Phi(f_i) \leqq \Phi(f) \leqq \sum_{i=1}^{3} \Phi(f_i), \text{ for any } f.$$

For, if $\Phi(f)$ is infinite then at least one $B(f_i) = \Phi(f_i)$ is infinite.

In conclusion, it may be of interest to look at certain peculiar examples of the Lebesgue area theory. The Geöcze surface of zero Lebesgue area and infinite Banach area is commonly cited as an example to show that the latter definition does not possess the lower semi-continuity character. It is clear that the sequence of polyhedra converging to the Geöcze surface which illustrated this point does not converge with respect to the distance function used here. A second surface of interest is one first given essentially by Nöbeling [9] and then used by Besicovitch [10] as an example of a homeomorphic image of the unit square in euclidean 3 space whose 3 dimensional Lebesgue measure exceeds its 2 dimensional Lebesgue area. The sequence of polyhedra converging to the given surface which shows this turns out also to be a fundamental sequence with respect to the distance $d(f, g)$. However, now, instead of converging to the continuous surface given by Besicovitch it converges to a discontinuous one. Actually $B(f) = \infty$ for the Besicovitch surface.

WAYNE UNIVERSITY.

BIBLIOGRAPHY.

[1] C. Goffman, "Lower semi continuity and area functionals, I. The non-parametric case," *Rend. Circ. Mat. di Palermo*, to appear.

[2] M. Fréchet, "Sur le prolongement des fonctionelles semi-continues et sur l'aire des surfaces courbes," *Fundamenta Mathematicae*, vol. 7 (1925), pp. 210-224.

[3] E. Silverman, "An intrinsic inequality for Lebesgue area," *Pacific Mathematical Journal*, to appear.

[4] T. Radó, *Length and Area*, New York, 1948, p. 438.

[5] J. L. Doob, *Stochastic Processes*, New York, 1953, p. 453.

[6] S. Banach, "Sur les lignes rectifiables et les surfaces dont l'aire est finie," *Fundamenta Mathematicae*, vol. 7 (1925), pp. 225-237.

[7] J. W. T. Youngs, "The topological theory of Fréchet surfaces," *Annals of Mathematics*, vol. 45 (1944), pp. 753-785.

[8] S. Saks, *Theory of the Integral*, Warsaw, 1937, p. 171.

[9] G. Nobeling, "Über der Flachenmasse in Euklidischen Raum," *Mathematische Annalen*, vol. 118 (1943), pp. 687-701.

[10] A. S. Besicovitch, "On the definition and the value of the area of a surface," *Quarterly Journal of Mathematics*, vol. 16 (1945), pp. 86-102.

ON THE ZEROS OF SOLUTIONS OF SECOND-ORDER LINEAR DIFFERENTIAL EQUATIONS.*[1]

By ZEEV NEHARI.

---

1. The present paper is concerned with the relation between the analytic function $p(z)$, defined in a domain D of the complex plane, and the location of the zeros of the solutions of the differential equation $w''(z) + p(z)w(z) = 0$. As in the case of a real variable, we shall refer to an equation as non-oscillatory if none of its solutions has an infinity of zeros on the set under discussion. If at least one solution has an infinity of zeros, the equation will be said to be oscillatory.

2. Our first result is a complex non-oscillation theorem for an infinite half-strip.

THEOREM I. *Let $p(z)$ be regular and let $|p(x+iy)| \leqq x\epsilon(x)$ in the half-strip $0 \leqq x < \infty$, $-1 < y < 1$ $(z = x+iy)$, where $\epsilon(x) \to 0$ for $x \to \infty$. If, for sufficiently small positive η,*

$$\int_0^\infty x\,|p[x \pm i(1-\eta)]|\,dx < \infty, \tag{1}$$

then the differential equation

$$w''(z) + p(z)w(z) = 0 \tag{2}$$

is non-oscillatory in any half-strip $0 \leqq x < \infty$, $-1+\delta \leqq y \leqq 1-\delta$ $(0 < \delta < 1)$.

Since $p(z)$ is regular for $-1+\delta \leqq y \leqq 1-\delta$, it is obvious that no solution of (2) can have an infinity of zeros on any of the linear segments $x = \text{const.}$, $-1+\delta \leqq y \leqq 1-\delta$. It is therefore sufficient to consider only zeros whose abscissas are different from each other.

If a and b $(\mathrm{Re}\{a\} < \mathrm{Re}\{b\})$ are two such zeros, we may transform (2) into an integral equation by means of the Green's function of the differential

---

* Received January 5, 1954.

[1] This research was supported by the United States Air Force through the Air Research and Development Command.

equation $w''(z) = 0$. Formally, the result is the same as in the real case, but the Green's function will now depend on the choice of an integration path C between a and b. We take C to be a rectifiable Jordan arc connecting a and b which is entirely contained in the half-strip in which $p(z)$ is regular. On C, we choose an arbitrary point ζ, and we denote by C_1 and C_2 the subarcs of C which connect a and ζ, and ζ and b, respectively. The Green's function in question is defined as follows.

$$(3)\quad g(z,\zeta) = (z-a)(b-\zeta)(b-a)^{-1},\ z \in C_1;$$
$$= (\zeta-a)(b-z)(b-a)^{-1},\ z \in C_2.$$

In view of the identity

$$[wg'-gw']_{C_1} + [wg'-gw']_{C_2} = \int_C (wg''-gw'')\,dz,\ g = g(z,\zeta),\ w = w(z),$$

it follows then that a solution of (2) which vanishes for $z = a$ and $z = b$ will satisfy the integral equation

$$(4)\qquad w(\zeta) = \int_C g(z,\zeta)\,p(z)\,w(z)\,dz.$$

We now identify ζ with the point (or one of the points) at which $|w(z)|$ attains its maximum on C. If M is this maximum, we conclude from (4) that $M \leqq \int_C |g(z,\zeta)|\,|p(z)|\,|w(z)|\,|dz| \leqq M\int_C |g(z,\zeta)|\,|p(z)|\,|dz|$, and therefore

$$(5)\qquad 1 \leqq \int_C |g(z,\zeta)|\,|p(z)|\,ds, \qquad |dz| = ds.$$

It may be remarked that (5) contains an inequality of Hartman and Wintner [3], to which it reduces if C is a segment of the real axis and $p(z)$ is real on C.

We next specialize the integration path C. If $a = a' + ia''$, $b = b' + ib''$ (a', a'', b', b'' real), we shall use a path consisting of three linear segments, and running from a to $a' + i(1-\delta)$, to $b' + i(1-\delta)$, to b.

By elementary geometry,

$$|(b-\zeta)/(b-a)| \leqq (1 + 4/(b'-a')^2)^{\frac{1}{2}} \leqq 1 + 2/(b'-a')^2 \qquad (\zeta \in C),$$

and we have the same bound for $|b-z|\,|b-a|^{-1}$, and also for $|\zeta-a|\,|z-a|^{-1}$ if $z \in C_2$. Hence, it follows from (3) that

$$|g(z,\zeta)| \leqq |z-a|(1+2/(b'-a')^2), \qquad z \in C_1,$$

$$|g(z,\zeta)| \leqq |z-a|(1+4/(b'-a')^2), \qquad z \in C_2.$$

If, as we shall assume, $a' \geqq 2$, we have $|z-a| \leqq x$, and thus

$$|g(z,\zeta)| \leqq x(1+4/(b'-a')^2), \qquad z \in C. \tag{6}$$

We shall, however, use (6) only on the horizontal part of C. On both vertical parts it follows from (3) by elementary considerations that $|g(z,\zeta)| \leqq 2(1+2/(b'-a')^2)$. Combining this and the inequality (6) with (5), and assuming that $b'-a' > 2$, we arrive at the inequality

$$1-4/(b'-a')^2 \leqq \int_{a'}^{b'} x\,|\,p[x+i(1-\delta)]\,|\,dx + 2\int_{-1+\delta}^{1-\delta} |\,p(a'+iy)|\,dy + 2\int_{-1+\delta}^{1-\delta} |\,p(b'+iy)|\,dy. \tag{7}$$

It is evident that we lost no generality by assuming that $a' \geqq 2$ and $b'-a' > 2$: if there is an infinity of zeros, we can choose a subset for which these assumptions are true. By the same argument, it is sufficient to consider subsets of zeros whose spacing is arbitrarily large. If $0 < \theta < 1$, we shall thus assume that for two consecutive zeros (the word "consecutive" referring to their real parts) a and b we have $b'-a' > 2(1-\theta)^{-\frac{1}{2}}$.

With this choice of a subset of zeros, the left-hand side of (7) reduces to θ. Suppose now that we have $n+1$ consecutive zeros $a_0, a_1, \cdots, a_n$, and add the n corresponding inequalities (7). This yields

$$n\theta \leqq \int_{a_0'}^{a_n'} x\,|\,p[x+i(1-\delta)]\,|\,dx + 4\sum_{\nu=0}^{n}\int_{-1+\delta}^{1-\delta} |\,p(a_\nu'+iy)|\,dy \tag{8}$$

(the right-hand side has been increased by doubling the first and last term of the summation). We now use the assumption (1). If

$$\int_0^\infty x\,|\,p[x+i(1-\delta)]\,|\,dx \leqq M,$$

we arrive at the inequality

$$n\theta \leqq M + 4\sum_{\nu=0}^{n}\int_{-1+\delta}^{1-\delta} |\,p(a_\nu'+iy)|\,dy. \tag{9}$$

We now have to estimate the integrals on the right-hand side of (9). To this end we consider the Cauchy integral

$$p(a) = \tfrac{1}{2\pi i}\int_\Gamma p(z)\,(z-a)^{-1}, \tag{10}$$

where the integration is extended over the perimeter of the rectangle $a'/2 - i(1-\delta/2)$, $2a' - i(1-\delta/2)$, $2a' + i(1-\delta/2)$, $a'/2 + i(1-\delta/2)$. On the upper horizontal boundary of the rectangle we have

$$|z-a|^{-1} = |x-a'+i(1-\delta/2-a'')|^{-1} \leqq |1-\delta/2-a''|^{-1} \leqq 2/\delta,$$

the last inequality following from $-1+\delta \leqq a'' \leqq 1-\delta$. Hence, if I_1 denotes the integral over this segment,

$$|I_1| \leqq (1/\pi\delta)\int_{a'/2}^{2a'} |p[x+i(1-\delta/2)]|\,dx$$
$$\leqq (2/\pi\delta a')\int_{a'/2}^{2a'} x\,|p[x+i(1-\delta/2)]|\,dx.$$

By (1), there exists a positive constant M_1 such that

$$\int_0^\infty x\,|p[x+i(1\pm\delta/2)]|\,dx < M_1. \tag{11}$$

Hence,

$$|I_1| \leqq 2M_1/\pi\delta a_1, \tag{12}$$

and this bound is also good for the contribution of the lower side of the rectangle to the integral (10). On the left side, we have

$$|z-a|^{-1} = |a'/2 - a' - i(y-a'')|^{-1} \leqq 2/a',$$

and on the right side

$$|z-a|^{-1} = |2a'-a'+i(y-a'')|^{-1} \leqq 1/a'.$$

Thus, by (10) and (12),

$$|p(a)| \leqq 4M_1/\pi\delta a' + (1/\pi a')\int_{-1+\delta/2}^{1-\delta/2} |p(a'/2+iy)|\,dy$$
$$+ 1/(2\pi a')\int_{-1+\delta/2}^{1-\delta/2} |p(2a'+iy)|\,dy. \tag{13}$$

By assumption, $|p(x+iy)| \leqq x\epsilon(x)$, where $\epsilon(x) \to 0$ for $x \to \infty$. It therefore follows from (13) that

$$|p(a)| \leqq 4M_1/\pi\delta a' + (1/\pi)\epsilon(a'/2) + (2/\pi)\epsilon(a').$$

Combining this with (9), we obtain

$$n\theta \leqq M + (32M_1/\pi\delta)\sum_{\nu=0}^{n} 1/a_\nu' + (8/\pi)\sum_{\nu=0}^{n}\epsilon(a_\nu'/2) + (16/\pi)\sum_{\nu=0}^{n}\epsilon(a_\nu').$$

For sufficiently large a_0', this inequality will be true *a fortiori* if $\epsilon(a_\nu'/2)$ and $\epsilon(a_\nu')$ are replaced by $\epsilon(a_0'/2)$. Writing, for short, $h(a_0') = (32M_1/\pi\delta a_0') + (24/\pi)\epsilon(a_0'/2)$, $M_2(a_0') = M + 32M_1/\pi\delta a_0' + (24/\pi)\epsilon(a_0'/2) \leqq M_3$, we thus have

$$n[\theta - h(a_0')] \leqq M_3, \tag{14}$$

where $h(a_0') \to 0$ for $a_0' \to \infty$, and the constant M_3 is independent of a_0'.

We now recall that the choice of the constant θ $(0 < \theta < 1)$ affected the spacing of the zeros of (2) to which our considerations refer, but did not influence the choice of the first zero, i. e., of a_0. If there were an infinity of zeros in our strip, we could identify a_0 with a zero whose abscissa is large enough so that $h(a_0') \leqq \theta/2$. It would then follow from (14) that $n \leqq 2M_3/\theta$, implying that the number of zeros with abscissas larger than a_0' is finite. The assumption that there is an infinite number of zeros in our half-strip thus leads to a contradiction, and Theorem I is proved.

3. In the proof of Theorem I, the assumption $|p(x+iy)| \leqq x\epsilon(x)$ played the role of a " Phragmén-Lindelöf " condition whose task was to ensure that the inequality (1) should imply $\lim p(x+iy) = 0$ as $x \to \infty$ throughout the half-strip. We may, instead, prescribe the asymptotic behavior of $|p(x+iy)|$ for large x and $-1 < y < 1$. While this procedure does not yield a non-oscillation condition, it leads to restrictions for the density of the zeros.

To illustrate the procedure, we assume $|p(x+iy)| < C/x^\sigma$, $C = \text{const.}$, $\sigma > 0$, and use (8). This yields

$$n\theta \leqq C_0 + C_1(a_n')^{2-\sigma} + C_2 \sum_{\nu=1}^{n} (a_\nu')^{-\sigma}, \tag{15}$$

where C_0, C_1, C_2, are suitable constants. Our aim is to find a type of asymptotic behavior of the a_n' which contradicts (15). Since $a_n' \to \infty$ for $n \to \infty$, the sum on the right-hand side of (15) can clearly be neglected in this connection. It follows that the assumption $(a_n')^{2-\sigma} = o(n)$ will contradict (16). However, before we formulate a result concerning the zeros of the differential equation, we have to consider the fact that $a_1, a_2, \cdots$ were not all the abscissas of the zeros. Since we are dealing with a subsequence for which $a_{\nu+1} - a_\nu$ is larger than a certain constant, the assumption that a_n' is $o(n^k)$, $0 < k \leqq 1$, will contradict (15) trivially, and no conclusion may therefore be drawn in this case with respect to the asymptotic distribution of the zeros.

THEOREM II. *If $p(z)$ is regular in the half-strip $0 \leqq x < \infty$, $-1 < y < 1$ $(z = x + iy)$ and if $a_1', a_2', \cdots (a_1' < a_2' < \cdots)$ are the abscissas of the zeros of any solution of the equation $w''(z) + p(z)w(z) = 0$ in the half-strip in question, then the assumption*

$$| p(x+iy) | \leqq C/x^{\sigma}, \quad 1 < \sigma < 2, \quad 0 \leqq x < \infty, \quad -1 < y < 1$$

is incompatible with $a_n' = o(n^{1/(2-\sigma)})$.

Theorem II is an analogue of a result of Hartman and Wintner [2] for the case of a real function $p(x)$ of the real variable x.

4. Our next result gives a non-oscillation criterion for the unit circle.

THEOREM III. *If $p(x)$ is regular in $| z | < 1$ and*

$$\int_0^{2\pi} | p(e^{i\theta}) | \, d\theta < \infty, \tag{16}$$

then the equation

$$w''(z) + p(z)w(z) = 0 \tag{17}$$

is non-oscillatory in $| z | < 1$.

We remark that the integral on the left-hand side of (16) is defined as the limit, for $\rho \to 1$, of the non-decreasing function $\int_0^{2\pi} | p(\rho e^{i\theta}) | \, d\theta$.

The proof of Theorem III will again be based on the inequality (5). If a and b ($| a | < 1, | b | < 1$) are zeros of a solution of (17), we take C to be the uniquely determined arc of the circle orthogonal to $| z | = 1$ which passes through a and b and is contained in $| z | < 1$. (3) shows that in this case (as generally in the case of a convex arc C) $| g(z, \zeta) | \leqq | g(\zeta, \zeta) |$, and we may therefore deduce from (5) that $1 \leqq \max_{\zeta \in C} | g(\zeta, \zeta) | \int_C | p(z) dz |$. Since $| b - \zeta | \leqq | b - a |$ and $| \zeta - a | \leqq | b - a |$, it follows from (3) that $| g(\zeta, \zeta) | \leqq | b - a |$. Hence,

$$1 \leqq | b - a | \int_C | p(z) dz |. \tag{18}$$

Our next step depends on the following result of Fejér and Riesz [1]: If $g(z)$ is regular in $| z | < 1$, then

$$\int_{-1}^{1} | g(x) | \, dx < \tfrac{1}{2} \int_0^{2\pi} | g(e^{i\theta}) | \, d\theta, \qquad z = x + iy. \tag{19}$$

If $g(z) = F'(z)$ and C_0 denotes the linear segment $-1 \leqq z \leqq 1$, this may be written in the form $\int_{C_0} | F'(z)dz | < \frac{1}{2} \int_{|z|=1} | F'(z)dz |$. Evidently, this inequality is unaffected by a transformation of the variable z. Hence, (19) implies the more general inequality

$$\int_{C'} | g(z)dz | < \tfrac{1}{2} \int_{\Gamma} | g(z)dz |, \tag{20}$$

where Γ is a rectifiable Jordan curve and C' is an arc which is transformed into a diameter of $| z | = 1$ by a conformal mapping of the interior of Γ onto the unit disk.

In particular, (20) will hold if Γ is the circumference $| z | = 1$ and C' is the intersection of $| z | \leqq 1$ and any circle orthogonal to $| z | = 1$, since a linear substitution $w = (z - \alpha)/(1 - \alpha^* z)$, $| \alpha | < 1$, which maps the unit disk onto itself will transform C' into a diameter if α is appropriately chosen. If C' belongs to the same orthogonal circle as does the integration path C in (18), we have $C \subset C'$ and therefore, by (20),

$$\int_C | p(z)dz | < \tfrac{1}{2} \int_0^{2\pi} | p(e^{i\theta}) | \, d\theta.$$

Combining this with (18), we obtain

$$2/| b - a | < \int_0^{2\pi} | p(e^{i\theta}) | \, d\theta. \tag{21}$$

It may be remarked that the constant 2 on the left-hand side of (21) can be easily improved to 4 by means of elementary geometric considerations; however, for our present purposes this is unnecessary.

Suppose now that there exists a solution of (17) which has an infinity of zeros in $| z | < 1$. If $a_1, a_2, \cdots$ is a converging infinite subset of these zeros, we have $\lim_{n \to \infty} | a_{n+1} - a_n | = 0$. Since, by (21),

$$2/| a_{n+1} - a_n | < \int_0^{2\pi} | p(e^{i\theta}) | \, d\theta,$$

it follows therefore that (16) is incompatible with the existence of an infinity of zeros in the unit disk. This completes the proof of Theorem III.

5. Theorem III is easily generalized to the case of a domain which is bounded by an analytic Jordan curve.

THEOREM IV. *If $p(z)$ is regular in a domain D bounded by an analytic Jordan curve C, and*

$$\int_C |\, p(z)dz \,| < \infty, \tag{22}$$

then the equation

$$w''(z) + p(z)w(z) = 0 \tag{23}$$

is non-oscillatory in D.

The integral in (22) may be defined as follows. We choose a point ζ of D and denote by C_m the level curve $g(z, \zeta) = m$ $(m > 0)$ of the harmonic Green's function $g(z, \zeta)$ of D. A conformal mapping if D onto the unit circle which carries ζ into the origin shows that $I(m) = \int_{C_m} |\, p(z)dz \,|$ grows monotonically with decreasing m. For $m \to 0$, C_m tends to C, and $\lim_{m\to 0} I(m)$ may thus be taken as a definition of the integral in (22). It is obvious that in the case in which D is the unit circle and ζ the origin, this definition coincides with the one used in Theorem III.

Let now $z = h(u)$ map D conformally onto $|\, u \,| < 1$, and rewrite equation (23) in terms of the variable u. This yields

$$v''(u) + q(u)v(u) = 0, \tag{24}$$

where $w[h(u)] = \sigma(u)v(u)$ and $\sigma(u)$ is a regular and non-vanishing function in $|\, u \,| < 1$. By classical results, $p(z)$ and $q(u)$ are connected by the identity

$$p(z)h'^2(u) = q(u) - \tfrac{1}{2}\{h(u), u\}, \tag{25}$$

where $\{h(u), u\} = (h''(u)/h'(u))' - \frac{1}{2}(h''(u)/h'(u))^2$ is the Schwarzian derivative of $h(u)$. Since C is an analytic curve and $h(u)$ maps the interior of C onto the unit circle, the function $h(u)$ is regular for $|\, u \,| \leqq 1$. Hence, $h'(u)$ is regular and $\neq 0$, and $\{h(u), u\}$ is regular, in $|\, u \,| \leqq 1$. By (25),

$$\lim_{m\to 0} \int_{C_m} |\, p(z)dz \,| = \lim_{\rho\to 1} \int_0^{2\pi} |\, q(u) - \tfrac{1}{2}\{h(u), u\} \,|\, |\, h'(u) \,|^{-1} d\theta,$$

where $u = \rho e^{i\theta}$, $0 < \rho < 1$, and it follows therefore that

$$\lim_{m\to 0} \int_{C_m} |\, p(z)dz \,| \tag{26}$$

and

$$\lim_{\rho\to 1} \int_0^{2\pi} |\, q(\rho e^{i\theta}) \,|\, d\theta \tag{27}$$

will be either both finite or both infinite.

By Theorem III, equation (24) will be non-oscillatory in $|u| < 1$ if (27) is finite. But since the solutions $v(u)$ of (24) and the solutions $w(z)$ of (23) are related by $w(z) = v(u)\sigma(u)$, where $\sigma(u)$ is regular and $\neq 0$ in $|u| < 1$, (23) and (24) are either both oscillatory or both non-oscillatory. Hence, (23) is non-oscillatory if (26) is finite. This proves Theorem IV.

WASHINGTON UNIVERSITY.

REFERENCES.

[1] L. Fejér and F. Riesz, "Über einige funktionentheoretische Ungleichungen," *Mathematische Zeitschrift*, vol. 11 (1921), pp. 305-314.

[2] P. Hartman and A. Wintner, "A criterion for the non-degeneracy of the wave equation," *American Journal of Mathematics*, vol. 71 (1949), pp. 206-213.

[3] ——— and A. Wintner, "On an oscillation criterion of Liapounoff," *American Journal of Mathematics*, vol. 73 (1951), pp. 885-890.

LIE ALGEBRA KERNELS AND COHOMOLOGY.*

By G. HOCHSCHILD.

Introduction. Let ϕ be a homomorphism of a Lie algebra E onto another Lie algebra L, and let M denote the kernel of ϕ. In this situation, we can associate with each element x of L the set $\psi(x)$ of derivations $m \to [e, m]$ of M, where e ranges over the subset $\phi^{-1}(x)$ of E. Then ψ is actually a homomorphism of L into the Lie algebra $D(M)/I(M)$, where $D(M)$ is the derivation algebra of M, and $I(M)$ is the ideal consisting of all the inner derivations $m \to [e, m]$, with $e \in M$. The pair (M, ψ) is called an extendible L-kernel. Generally, an L-kernel is a pair (M, ψ) consisting of an arbitrary Lie algebra M and a homomorphism ψ of L into $D(M)/I(M)$. Not every L-kernel is extendible, i. e., not every L-kernel arises, in the way explained above, from an extension (E, ϕ) by L. A criterion for extendibility of a kernel can be given in terms of cohomology theory. With every L-kernel, one can associate a certain 3-dimensional cohomology class for L in the center of M, and the L-kernel is extendible if and only if this cohomology class is 0. This is the precise analogue of a result of Eilenberg and MacLane [1], for the case of groups, and can easily be proved in the same way, mutatis mutandis.

Pursuing the analogy with the theory of Eilenberg and MacLane a little further, one is led to consider equivalence classes of kernels, two kernels being equivalent if (as L-modules) their centers coincide and if, furthermore, they differ (in a sense to be made precise later) only by an extendible kernel. One can then show that the set of equivalence classes of kernels, for a fixed Lie algebra L, and a fixed L-module N as center, can be made into a vector space in such a way that the correspondence between kernels and cohomology classes induces a linear isomorphism of the space of the kernel classes into that of the 3-dimensional cohomology group $H^3(L, N)$.

The first difficulty arises when one tries to show that this map is *onto*, i. e., that every element of $H^3(L, N)$ arises from an L-kernel with center N. This requires the construction of an L-kernel from a given 3-cocycle for L in N. It appears that this is difficult to accomplish directly. (The case of a semisimple Lie algebra of characteristic 0 shows that one must abandon finite dimensionality in such a construction.) However, if one employs the uni-

* Received February 2, 1954.

versal enveloping algebra of L one obtains cocycles of the associative type as representatives for the elements of $H^3(L, N)$, and with these one can construct corresponding L-kernels by a slight modification (actually a simplification) of the construction of associative kernels which we used in [2], where we dealt with the analogous theory for associative algebras.

The corresponding theory for *restricted* Lie algebras over fields of characteristic p is considerably more complicated. Here, the enveloping algebra ceases to be purely auxiliary but plays a principal part. Also, contrary to the case of ordinary Lie algebras, the passage from a *restricted* kernel to a *restricted* cohomology class (which cannot be defined without reference to the enveloping algebra) is even more laborious than the passage from a (restricted) cohomology class to a (restricted) kernel. For this reason, our procedure below is in the direction opposite to that of our sketch above. The final result is that the space of the restricted kernel classes is isomorphic with that of the restricted 3-dimensional cohomology group $H^3{}_*(L, N)$, where N is a restricted L-module with $N^{[p]} = (0)$. For our terminology and for the elementary results concerning restricted Lie algebras and restricted cohomology, we refer the reader to [3]. Also, one of the main results of that paper is used in section 2.

It goes without saying that the case of ordinary Lie algebras is implicitly contained in that of restricted Lie algebras. One merely omits everything relating to the p-map. Note, however, that our above account of the ordinary case ties up with our remarks at the end of section 1 to show that section 3 is completely superfluous in the case of ordinary Lie algebras.

1. Restricted kernels. Let M be a restricted Lie algebra over a field F of characteristic p. If m is an element of M we denote by D_m the inner derivation of M which is effected by m, i. e., $D_m(u) = [m, u]$. A derivation ζ of M is said to be a restricted derivation if $\zeta(m^{[p]}) = D_m{}^{p-1}(\zeta(m))$, for all $m \in M$. In particular, every inner derivation of M is restricted. The restricted derivations of M constitute a restricted Lie algebra, with the commutation $[\zeta, \tau] = \zeta\tau - \tau\zeta$ and the p-map $\zeta^{[p]} = \zeta^p$. We denote this restricted Lie algebra of the restricted derivations of M by $D_0(M)$. The inner derivations of M constitute a restricted ideal $I(M)$ of $D_0(M)$, so that we have a restricted factor Lie algebra $D_0(M)/I(M)$.

Let L be another restricted Lie algebra over F. The structure of a *restricted L-kernel* on M is defined by giving a restricted homomorphism ψ of L into $D_0(M)/I(M)$. This homomorphism ψ defines the structure of a restricted L-kernel on the center N of M. This restricted L-kernel N is called

the *nucleus* of the restricted L-kernel M. Note that, since N is abelian, the ψ-structure of N is the structure of a restricted L-module, while the restricted Lie algebra structure of N is reduced simply to a p-semilinear map $n \to n^{[p]}$ of N into the annihilator N^L of L in N. When we refer to a nucleus we intend this to comprise both of these structures.

Let M_1 and M_2 be two restricted L-kernels with a common nucleus N. We construct a composite restricted L-kernel M with nucleus N as follows. In the direct sum (M_1, M_2) of the restricted Lie algebras M_1 and M_2, let J be the restricted ideal consisting of the pairs $(n, -n)$ with $n \in N$. We take $M = (M_1, M_2)/J$, where we identify the center $(N, N)/J$ with N by means of the isomorphism which is induced by the homorphism $(n_1, n_2) \to n_1 + n_2$ of (N, N) onto N. Clearly, M_1 and M_2 may be regarded as restricted ideals of M, and when this is done we have $M = M_1 + M_2$, $M_1 \cap M_2 = N$, and $[M_1, M_2] = (0)$. The L-kernel structure of M is defined from the given restricted homomorphisms ψ_i of L into $D_0(M_i)/I(M_i)$ as follows. If $x \in L$, let ζ_i be a restricted derivation of M_i which belongs to the class $\psi_i(x)$. There is a unique restricted derivation ζ of M which coincides with ζ_i on M_i, for $i = 1, 2$. The coset mod. $I(M)$ of ζ is evidently determined uniquely by $\psi_1(x)$ and $\psi_2(x)$, and we denote it by $\psi(x)$. Then ψ is a restricted homomorphism of L into $D_0(M)/I(M)$ and defines the structure of a restricted L-kernel on M. Clearly, the nucleus of M is N.

We may evidently identify $M_1 + M_2$ with $M_2 + M_1$, and $M_1 + (M_2 + M_3)$ with $(M_1 + M_2) + M_3$. The restricted L-kernels with a common nucleus N do not constitute an abelian group with this composition. However, one obtains the structure of an abelian group in the set of certain equivalence classes of such kernels. The equivalence relation which we shall introduce is based on the point of view from which extendible kernels are disregarded. Here, a restricted L-kernel M is said to be extendible if there exists a restricted extension (E, ϕ) of M by L which induces the given L-kernel structure on M, i. e., which is such that, for every $e \in E$, the restriction of D_e to M belongs to the class $\psi(\phi(e))$ of restricted derivations of M.

If M_1 and M_2 are restricted L-kernels with a common nucleus N we say that M_1 and M_2 are equivalent if there exist extendible restricted L-kernels U and V with nucleus N such that $M_1 + U = M_2 + V$. This relation is evidently symmetric and reflexive. The transitivity follows at once from the fact that a composite of extendible kernels is extendible. Indeed, if (E_1, ϕ_1) and (E_2, ϕ_2) are restricted extensions by L whose kernels M_1 and M_2 have the same nucleus N, we can consider the restricted subalgebra of the direct sum (E_1, E_2) which consists of all pairs (e_1, e_2) in which $\phi_1(e_1) = \phi_2(e_2)$.

If we reduce this restricted Lie algebra by the restricted ideal of the elements $(n, -n)$, with $n \in N$, we obtain a restricted Lie algebra E containing $M_1 + M_2$ as a restricted ideal. There is an evident restricted homomorphism ϕ of E onto L (induced by ϕ_1 or ϕ_2) such that (E, ϕ) is a restricted extension of $M_1 + M_2$ by L which induces the proper kernel structure on $M_1 + M_2$, whence $M_1 + M_2$ is extendible.

We shall see that the equivalence classes of the restricted L-kernels with a fixed nucleus N constitute an abelian group, with the composition of kernel classes that is induced by the composition of kernels as defined above. Moreover, in the case where $N^{[p]} = (0)$ (which is the case of principal interest here), we can even define the structure of a vector space over the basic field F on this group of kernel classes. This is done as follows.

Let M be a restricted L-kernel with nucleus N, and let $\alpha \in F$, $\alpha \neq 0$. Denote by J_α the ideal of the direct sum (M, N) which consists of the elements $(n, -\alpha n)$, with $n \in N$. If $N^{[p]} = (0)$ we have $J_\alpha{}^{[p]} = (0)$ so that J_α is, in particular, a restricted ideal of (M, N). In the general case, it is still true that J_α is a restricted ideal of (M, N) whenever α belongs to the prime field of F, and in particular for $\alpha = -1$. Now we form $(M, N)/J_\alpha$ and identify its center $(N, N)/J_\alpha$ with N by means of the isomorphism which is induced by the homomorphism $(n_1, n_2) \to \alpha n_1 + n_2$ of (N, N) onto N. The L-kernel structure of M induces on $(M, N)/J_\alpha$ the structure of a restricted L-kernel with nucleus N. This kernel will be denoted αM. It is easy to see that $1M$ may be identified with M, namely by the map which is induced by the map $(m, n) \to m + n$. Also, $\alpha(\beta M)$ may be identified with $(\alpha\beta)M$. In fact, the elements of $\alpha(\beta M)$ are represented in (M, N, N) by triples (m, n_1, n_2), and it is easily seen that the map $(m, n_1, n_2) \to (m, \alpha n_1 + n_2)$ induces the required isomorphism of $\alpha(\beta M)$ onto $(\alpha\beta)M$. Furthermore, one checks almost immediately that $\alpha(M_1 + M_2)$ may be identified with $\alpha M_1 + \alpha M_2$. Observing that the passage $M \to \alpha M$ can be carried over to restricted extensions of M by L, we may conclude that αM is extendible whenever M is extendible. It follows that the maps $M \to \alpha M$ induce corresponding maps in the set of equivalence classes of kernels.

It is quite possible to give non-computational proofs for the fact that the equivalence classes of kernels constitute an abelian group with the composition defined above, and that our maps $M \to \alpha M$ induce automorphisms of the group of kernel classes that represent the multiplicative group of F in the automorphism group of the group of kernel classes. The only difficulty resides in the treatment of the additive structure of F. Note that we have not even defined the zero multiple of a kernel. The above definition fails

for $\alpha = 0$, because the center of $(M, N)/J_0$ may very well be larger than $(N, N)/J_0$.

Before settling these questions concerning the group structure and the vector space structure (in case $N^{[p]} = (0)$) of the set of kernel classes, let us enumerate the facts we have to establish. The 0-element of the group of the kernel classes will be the class consisting of the extendible kernels. This means that if M and U are restricted L-kernels with a common nucleus such that U and $M + U$ are extendible then M is extendible. The negative of the equivalence class of M is the equivalence class of $(-1)M$. This means that the kernel $M + (-1)M$ is extendible. Finally, if M is a restricted L-kernel with nucleus N such that $N^{[p]} = (0)$, and if α, β, and $\alpha + \beta$ are non-zero elements of M then $\alpha M + \beta M + (-\alpha - \beta)M$ is extendible.

All these facts will follow easily from a criterion for extendibility of a kernel which we shall establish next. Let (M, ψ) be a restricted L-kernel with an arbitrary nucleus N. We can find a linear map $x \to t_x$ of L into $D_0(M)$, called an *operator map*, such that t_x belongs to the class $\psi(x)$, for every $x \in L$. We have then $[t_x, t_y] - t_{[x,y]} \in I(M)$, so that we can find an alternating bilinear map τ from L to M, called a *hindrance*, such that $[t_x, t_y] = t_{[x,y]} + D_{\tau(x,y)}$, for all x and y in L. Since ψ is a restricted homomorphism, we have $t_x{}^p - t_{x^{[p]}} \in I(M)$, and we choose an element x^* in M such that $t_x{}^p = t_{x^{[p]}} + D_{x^*}$. It is easily verified by induction on k that $D_{t_x}{}^k(t_y) = t_{y_k} + D_m$, where $y_k = D_x{}^k(y)$, and $m = \sum_{i=0}^{k-1} t_x{}^i(\tau(x, y_{k-1-i}))$. It follows at once from this that $[D_{x^*}, t_y]$ is equal to the inner derivation of M which is effected by the element $-\tau(x^{[p]}, y) + \sum_{i=0}^{p-1} t_x{}^i(\tau(x, y_{p-1-i}))$. This means that the element $T_x(y) = t_y(x^*) - \tau(x^{[p]}, y) + \sum_{i=0}^{p-1} t_x{}^i(\tau(x, D_x{}^{p-1-i}(y)))$ belongs to N. We shall say that T (or sometimes $T_x(y)$) is a *secondary deviation* of M.

We obtain what we shall call a *primary deviation* of M by taking the formal coboundary of τ, i. e., the alternating trilinear map $\delta\tau$ defined on L by the formula

$$\begin{aligned}\delta\tau(x, y, z) = {} & t_x(\tau(y, z)) - t_y(\tau(x, z)) + t_z(\tau(x, y)) \\ & - \tau([x, y], z) + \tau([x, z], y) - \tau([y, z], x).\end{aligned}$$

A straightforward verification shows that $\delta\tau$ takes its values in the center N of M, and that it is actually a 3-cocycle for L in N whose cohomology class in $H^3(L, N)$ depends only on (M, ψ), and not on the particular choice of t and τ (The proof is given in [4]).

If M is extendible, we can choose the operator map t, the hindrance τ and the elements x^* in such a way that the deviations become 0. In fact, let (E, ϕ) be a restricted extension of M by L which induces the given L-kernel structure on M. Then we can find a linear map $x \to e(x)$ of L into E such that $\phi(e(x)) = x$, for all $x \in L$. We may then define $t_x(m) = [e(x), m]$, for $x \in L$ and $m \in M$. For a corresponding hindrance, we may take $\tau(x, y) = [e(x), e(y)] - e([x, y])$. We may take $x^* = e(x)^{[p]} - e(x^{[p]})$, and it is easy to verify that we then obtain $T_x(y) = 0$. Also, a direct computation shows that $\delta\tau = 0$. Conversely, we have the following result.

LEMMA 1.1. *Let M be a restricted L-kernel, and suppose it is possible to choose the operator map t, the hindrance τ, and the elements x^*, corresponding to the elements x of a basis for L over F, in such a way that $\delta\tau = 0$, and $T_x(y) = 0$, for all $y \in L$ and each basis element x. Then M is extendible.*

Proof. Since $\delta\tau = 0$, we can define an extension (E, ϕ) of M by L as follows. As a vector space, E is the direct sum (L, M) of L and M. The commutation in E is defined by the formula

$$[(x, u), (y, v)] = ([x, y], t_x(v) - t_y(u) + [u, v] + \tau(x, y)).$$

Indeed, the Jacobi identity is equivalent to $\delta\tau = 0$. The homomorphism ϕ of E onto L is defined by $\phi(x, u) = x$. There remains to define a p-map in E such that (E, ϕ) becomes a restricted extension of M by L which induces the given kernel structure on M. Let R_E denote the universal enveloping algebra of E, and let (x_i) be the basis for L over F for which the x_i^* have been defined. We write a given element x of L in the form $x = \sum_i \alpha_i x_i$. If $u \in M$ we define $(x, u)^{[p]}$ within R_E by the following formula, in which it is assumed that R_M has been identified with a subalgebra of R_E in the canonical fashion: $(x, u)^{[p]} = (x, u)^p - (u^p - u^{[p]}) - \sum_i \alpha_i^p((x_i, 0)^p - (x_i^{[p]}, x_i^*))$. By the usual computation with p-th powers in an associative algebra (all the formal results of this kind which are needed are collected in [3]), one sees that this element is actually the sum of a linear combination of commutators of elements of E and the element $u^{[p]} + \sum_i \alpha_i^p(x_i^{[p]}, x_i^*)$, so that it actually lies in E. The map $(x, u) \to (x, u)^{[p]}$ has the additive property of a p-map in virtue of our definition, and it is also evident that $(\alpha(x, u))^{[p]} = \alpha^p(x, u)^{[p]}$. Hence, in order to conclude that it is actually a p-map, it suffices to show that $(x, u)^p - (x, u)^{[p]}$ is in the center of R_E. Now $u^p - u^{[p]}$ is in the center of R_M. Also, since t_y is a restricted derivation, $(y, 0)(u^p - u^{[p]}) = (u^p - u^{[p]})(y, 0)$, for every $y \in L$. Hence $u^p - u^{[p]}$ is in the center of R_E.

Hence it suffices to show that each $(x_i, 0)^p - (x_i^{[p]}, x_i^*)$ lies in the center of R_E. Since $t_{x_i^p} - t_{x_i^{[p]}} = D_{x_i \bullet}$, each of these elements commutes with every element of M. On the other hand, the computation we made above in connection with $T_x(y)$ shows that

$$((x_i, 0)^p - (x_i^{[p]}, x_i^*))(y, 0) - (y, 0)((x_i, 0)^p - (x_i^{[p]}, x_i^*)) = (0, T_{x_i}(y)) = 0.$$

Hence we have indeed a p-map with which (E, ϕ) is a restricted extension of M by L. It is evident that this extension induces the given kernel structure on M. This completes the proof of Lemma 1. 1.

Now we observe that we can carry deviations along with our operations on kernels in a linear fashion. More precisely, if M_1 and M_2 are two restricted L-kernels with a common nucleus and if we have deviations for M_1 and M_2, the natural carrying over of the operators, hindrances, and elements x^* will yield deviations for $M_1 + M_2$ which are the sums of the deviations for M_1 and M_2. Similarly, the α-multiples of deviations for M are deviations for αM. These facts can be seen directly from our definitions, and require no computational proofs. It is clear how one can now apply Lemma 1. 1 in order to show that our operations on kernels induce the structure of a group or, if $N^{[p]} = (0)$, of a vector space in the set of equivalence classes of restricted L-kernels with a common nucleus N.

As we have already remarked above, the primary deviations of a restricted L-kernel M with nucleus N all belong to a single cohomology class in $H^3(L, N)$. Furthermore, it follows from what we have just seen that this yields a homomorphism of the group of the kernel classes into $H^3(L, N)$ which is, moreover, a linear map in the case where $N^{[p]} = (0)$. If everything concerning the p-maps is ignored, and ordinary kernels for ordinary Lie algebras are considered, it is easy to see that the linear map of the space of the kernel classes into $H^3(L, N)$ is an isomorphism. In fact, if a kernel has a primary deviation $\delta\tau$ which is a coboundary, i. e., of the form $\delta\sigma$, where σ is a 2-cochain for L in N, then we can replace the hindrance τ by the hindrance $\tau - \sigma$, and thus obtain 0 as a primary deviation. By Lemma 1. 1, it then follows that the kernel is extendible. Finally, in the case of ordinary Lie algebras, the construction of kernels which we shall give in the next section leads at once to the conclusion that every cohomology class in $H^3(L, N)$ arises from an L-kernel with nucleus N, so that $H^3(L, N)$ is isomorphic with the space of the kernel classes, by the linear map which is obtained from the primary deviations, just as in [1] for the case of groups.

2. Construction of kernels. We begin with the construction of an associative kernel for an arbitrary associative algebra U over the field F.

Then we shall take for U the u-algebra of a restricted Lie algebra L and turn the U-kernel into a restricted L-kernel by replacing the multiplications with commutations.

We denote by $u \to u'$ a linear isomorphism of U onto a copy U' of the underlying vector space of U. Let I denote the sum of the homogeneous components of degree 1 and 2 of the tensor space over U', relative to F, i. e., $I = U' + U' \otimes U'$. We introduce an auxiliary associative algebra C of dimension 2 over F: $C = Fe + Ff$, where $e^2 = e$, $f^2 = f$, $ef = f$, $fe = e$. We make the direct sum $J = C + I$ into an associative algebra over F, containing C as a subalgebra, and such that $IJ = (0)$, while $ev = v = fv$, for every $v \in I$. Now let K denote the direct sum $J + U \otimes U$. It is easy to verify that there is a multiplication in K with which K is an associative algebra over F containing J as an ideal, and such that the following requirements (which evidently characterize the multiplication) are satisfied:

$(U' \otimes U' + U \otimes U)K = (0)$; for u, u_1, and u_2 in U,

$e(u_1 \otimes u_2) = u_1' \otimes u_2'$, $f(u_1 \otimes u_2) = 0$, and

$$u'(u_1 \otimes u_2) = (uu_1)' \otimes u_2' - u' \otimes (u_1u_2)'.$$

One can verify directly that the center of K is (0). Finally, let N be a left U-module over F, and put $M = K + N$. We make M into an associative algebra over F such that $MN = (0) = NM$. Then N is the center of M.

Now let g be a given 3-dimensional cocycle for U in N. This means that g is a trilinear map from U to N satisfying the identity $x \cdot g(y, z, t) - g(xy, z, t) + g(x, yz, t) - g(x, y, zt) = 0$. We can make M into a left U-module containing N as a U-submodule, and such that $U \cdot J = (0)$ and $u \cdot (u_1 \otimes u_2) = uu_1 \otimes u_2 - u \otimes u_1u_2 + g(u, u_1, u_2)$. We claim that $u \cdot (ab) = (u \cdot a)b$, for all a and b in M and all u in U. Indeed, since J is an ideal in M which is annihilated by U, $u \cdot (ab)$ and $(u \cdot a)b$ are both 0 if $a \in J$. If $a \in U \otimes U + N$ then $ab = 0$ and $u \cdot a \in U \otimes U + N$, so that $(u \cdot a)b = 0$. This suffices in order to establish our claim.

Next we define right U-operators in M such that

$$(U' \otimes U' + U \otimes U + N) \cdot U = (0),\ e \cdot u = u',\ f \cdot u = 0, \text{ and}$$

$$u_1' \cdot u = (u_1u)' + u_1' \otimes u'.$$

These right U-operators commute with the left U-operators. On J, this follows at once from the fact $J \cdot U \subset J$ and $U \cdot J = (0)$. On $U \otimes U + N$, this follows from the fact that $U \cdot (U \otimes U + N) \subset U \otimes U + N$ and $(U \otimes U + N) \cdot U = (0)$.

Furthermore, $(ab)\cdot u = a(b\cdot u)$, for all a and b in M and all u in U. Indeed, if b lies in $U'\otimes U' + U\otimes U + N$ then so does ab, whence $(ab)\cdot u = 0 = a(b\cdot u)$. Hence we may assume $b \in C + U'$. Then, if $a \in I + U\otimes U + N$, $ab = 0$ and $b\cdot u \in I$ so that $a(b\cdot u) = 0$. Hence we may asume that $a \in C$ and $b \in C + U'$, and in this case it is easy to verify directly that $(ab)\cdot u = a(b\cdot u)$.

Finally, we claim that $(a\cdot u)b = a(u\cdot b)$. Indeed, if $a \in U\otimes U + N$ we have $(a\cdot u)b = 0 = a(u\cdot b)$. Hence we may suppose that $a \in J$. Then $a\cdot u \in I$. Now if $b \in J + N$ we have $u\cdot b \in N$ and therefore $(a\cdot u)b = 0 = a(u\cdot b)$. Hence we may now assume that $a \in J$ and $b \in U\otimes U$. Now if $a \in Ff + U'\otimes U'$ we have $a\cdot u = 0$ and $a(u\cdot b) = 0$. There remains the case $a \in Fe + U'$ and $b \in U\otimes U$. In this case, the relation $(a\cdot u)b = a(u\cdot b)$ is easily verified directly.

With our right U-operators, M is *not* a right U-module. Instead of $(a\cdot u_1)\cdot u_2 = a\cdot(u_1u_2)$, we have actually $(a\cdot u_1)\cdot u_2 - a\cdot(u_1u_2) = a(u_1\otimes u_2)$. Indeed, each side is 0 whenever $a \in Ff + U'\otimes U' + U\otimes U + N$. The remaining case $a \in Fe + U'$ is easily checked directly.

Now take $U = U_L{}^+$, where L is a restricted Lie algebra over F. ($U_L = R_L/P_L$, where P_L is the ideal generated by the elements of the form $x^p - x^{[p]}$, with $x \in L$. $U_L{}^+$ is the ideal of U_L which is generated by the elements of L.) Then N is already defined as a restricted L-module, and the given cocycle g represents an element of the restricted cohomology group $H_*{}^3(L, N)$. We regard M as a restricted Lie algebra with $[a, b] = ab - ba$ and $a^{[p]} = a^p$. If $x \in L$ we denote by t_x the *restricted* derivation of M which is defined by $t_x(a) = x\cdot a - a\cdot x$ (working in the ring of linear transformations of $Fx \oplus M$, one sees at once that $t_x(a^{[p]}) = D_{a^{p-1}}(t_x(a))$). We have then

$$(t_xt_y - t_yt_x - t_{[x,y]})(a)$$
$$= (a\cdot y)\cdot x - a\cdot(yx) - (a\cdot x)\cdot y + a\cdot(xy) = [x\otimes y - y\otimes x, a].$$

Hence t induces a homomorphism of L into $D_0(M)/I(M)$. If u_x denotes the left x-operator on M and v_x the right x-operator we have $t_x = u_x - v_x$. Since u_x commutes with v_x, we have therefore $t_x{}^p = u_x{}^p - v_x{}^p$. Since M is a left $U_L{}^+$-module, $u_x{}^p = u_{x^{[p]}}$. On the other hand, one finds by induction on k that, for $k > 1$, $v_x{}^k(a) = a\cdot x^k + a(x^{k-1}\otimes x)$, and so $v_x{}^p = v_{x^{[p]}} - D_{x^*}$, where $x^* = x^{p-1}\otimes x$. Thus $t_x{}^p - t_{x^{[p]}} = D_{x^*}$, which shows that our homomorphism of L into $D_0(M)/I(M)$ is a restricted homomorphism, so that we have the structure of a restricted L-kernel on M. In order to exhibit the dependence of this structure on the given cocycle g, we shall now denote our kernel by M_g. Observe that the nucleus of M_g is N, with $N^{[p]} = N^p = (0)$.

If h is any 2-cochain for $U_L{}^+$ in N it is easily seen that $M_{\delta h}$ is extendible. In fact, we can actually define the structure of an associative algebra on the space $U_L{}^+ + M_{\delta h}$ which contains the associative algebra $M_{\delta h}$ as an ideal. The multiplication is defined by the formula

$$(u + a)(v + b) = uv + u \cdot b + a \cdot v + ab + u \otimes v - h(u, v),$$

where u and v are in $U_L{}^+$, and a and b in $M_{\delta h}$. We make this associative algebra into a restricted Lie algebra in the same way as we made M into a restricted Lie algebra. Then $L + M_{\delta h}$ is a restricted subalgebra, and if we define $\phi(x + a) = x$ the pair $(L + M_{\delta h}, \phi)$ is a restricted extension of $M_{\delta h}$ by L which induces the given kernel structure on $M_{\delta h}$.

Now we can easily find deviations for M_g which depend linearly on g. In fact, it is clear from the above that we may take $\tau(x, y) = x \otimes y - y \otimes x$ for the hindrance. A direct computation shows that the corresponding primary deviation $\delta\tau$ is given by $\delta\tau(x, y, z) = g(x, y, z) - g(x, z, y) + g(z, x, y) - g(z, y, x) + g(y, z, x) - g(y, x, z)$.

In order to compute the secondary deviation, write y_k for $D_x{}^k(y)$, and note that

$$t_x{}^i(\tau(x, y_{p-1-i})) = t_x{}^i(x \otimes y_{p-1-i} - y_{p-1-i} \otimes x) = x^i \cdot (x \otimes y_{p-1-i} - y_{p-1-i} \otimes x)$$
$$= x^{i+1} \otimes y_{p-1-i} - x^i \otimes y_{p-i} - x^i y_{p-1-i} \otimes x + g(x^i, x, y_{p-1-i}) - g(x^i, y_{p-1-i}, x),$$

for $i \geqq 1$. If we sum these terms from $i = 1$ to $i = p - 1$, and then add the term $x \otimes y_{p-1} - y_{p-1} \otimes x$, we find that

$$\sum_{i=0}^{p-1} t_x{}^i(\tau(x, y_{p-1-i})) = \sum_{i=1}^{p-1} (g(x^i, x, y_{p-1-i}) - g(x^i, y_{p-1-i}, x)) + x^{[p]} \otimes y$$
$$- \sum_{i=0}^{p-1} x^i y_{p-1-i} \otimes x.$$

By a well known identity, valid in any associative algebra of characteristic p, we have $\sum_{i=0}^{p-1} x^i y_{p-1-i} = y x^{p-1}$. If we substitute this above, we find that, using $x^* = x^{p-1} \otimes x$ in $T_x(y)$, we have

$$T_x(y) = g(y, x^{p-1}, x) + \sum_{i=1}^{p-1} (g(x^i, x, y_{p-1-i}) - g(x^i, y_{p-1-i}, x)).$$

By what we have seen in section 1, it follows from the linear dependence of these deviations on g that the map which sends g into the equivalence class of M_g is linear. Since $M_{\delta h}$ is extendible, this linear map induces a linear map of $H_*{}^3(L, N)$ into the space of the kernel classes with nucleus N, where $N^{[p]} = (0)$.

We shall show next that this linear map is an isomorphism. This means that if M_g is extendible then g is a coboundary. Suppose that (E, ϕ) is a restricted extension of M_g by L which induces the given kernel structure on M_g. Then we can find a linear map $x \to e(x)$ of L into E such that $\phi(e(x)) = x$ and $t_x(m) = [e(x), m]$, for all $x \in L$ and all $m \in M_g$. Then, for x and y in L, $[e(x), e(y)] - e([x, y])$ is an element of M_g, and the inner derivation which is effected by this element coincides with $[t_x, t_y] - t_{[x,y]}$, which is the inner derivation effected by $x \otimes y - y \otimes x$. Hence we have $[e(x), e(y)] - e([x, y]) - x \otimes y + y \otimes x \in N$. Similarly, the inner derivation which is effected by the element $e(x)^{[p]} - e(x^{[p]})$ coincides with $t_x{}^p - t_{x^{[p]}}$, which is the inner derivation effected by $x^{p-1} \otimes x$. Hence we have

$$e(x)^{[p]} - e(x^{[p]}) - x^{p-1} \otimes x \in N.$$

The component J of M_g is easily seen to be a restricted ideal of E, and M_g/J may be identified with the component $Q = U_L{}^+ \otimes U_L{}^+ + N$ of M_g. Put $S = E/J$, and let ϕ' be the homomorphism of S onto L which is induced by ϕ. Then (S, ϕ') is a restricted extension of Q by L. If $s(x)$ denotes the coset mod. J of $e(x)$ we have, from the above, $[s(x), s(y)] - s([x, y]) - x \otimes y + y \otimes x \in N$, and $s(x)^{[p]} - s(x^{[p]}) - x^{p-1} \otimes x \in N$. Since $[Q, Q] = (0) = Q^{[p]}$, we can apply the theory of [3] to (S, ϕ'). In particular, this theory shows that there is a 2-cocycle h for $U_L{}^+$ in Q such that (S, ϕ') is equivalent to the restricted extension (E_h, ϕ_h) which is constructed in the standard fashion from h, i. e., the space of E_h is the direct sum (L, Q), the commutation is given by the formula

$$[(x_1, q_1), (x_2, q_2)] = ([x_1, x_2], x_1 \cdot q_1 - x_2 \cdot q_2 + h(x_1, x_2) - h(x_2, x_1)),$$

the p-map is given by $(x, q)^{[p]} = (x^{[p]}, x^{p-1} \cdot q + h(x^{p-1}, x))$, and $\phi_h(x, q) = x$.

Now let us consider the extensions $(S/N, \phi^*)$ and $(E_h/N, \phi_h{}^*)$ which are obtained by reducing the above extensions mod. N. These are equivalent restricted extensions of Q/N by L. From our above relations involving the map s of L into S, it is clear that $(S/N, \phi^*)$ is equivalent to the extension (E_{k^*}, ϕ_{k^*}), where k is the 2-cochain for $U_L{}^+$ in Q which is defined by $k(u, v) = u \otimes v$, and k^* is the *2-cocycle* (for $\delta k = g$) for $U_L{}^+$ in Q/N which is obtained by reducing k mod. N. On the other hand, $(E_h/N, \phi_h{}^*)$ is equivalent to (E_{h^*}, ϕ_{h^*}), where h^* is obtained by reducing h mod. N. It follows from the theory of restricted extensions that there is a 1-cochain c^* for $U_L{}^+$ in Q/N such that $h^* + \delta c^* = k^*$. Choose a 1-cochain c for $U_L{}^+$ in Q which reduces to c^* mod. N. Then the 2-cochain $k - h - \delta c$ takes values in N, and $\delta(k - h - \delta c) = \delta k = g$, i. e., g is a coboundary. Hence we may

conclude that the map $g \to M_g$ induces a linear isomorphism of $H_*^3(L, N)$ into the space of the restricted kernel classes with nucleus N, and $N^{[p]} = (0)$. There remains to show that this actually is an isomorphism *onto*, and this is the purpose of the next section.

3. Construction of cocycles. We require a construction which attaches a 3-cocycle for U_L^+ in N to a given restricted L-kernel M with nucleus N, such that $N^{[p]} = (0)$. Let $x \to t_x$ be a linear map of L into $D_0(M)$ which represents the kernel homomorphism of L into $D_0(M)/I(M)$, and let τ be a corresponding hindrance, so that $[t_x, t_y] - t_{[x,y]} = D_{\tau(x,y)}$. The restricted derivations t_x of M extend uniquely to derivations of U_M which we still denote by t_x. We have then, for all x and y in L and all $u \in U_M$, $(t_x t_y - t_y t_x - t_{[x,y]})(u) = \tau(x, y)u - u\tau(x, y)$.

We shall construct a non-associative algebra Q over the tensor product (relative to the base field F) $U_L \otimes U_M$ which contains U_M as a subalgebra and whose deviation from associativity will yield the required 3-cocycle for U_L^+ in N. The underlying idea of the construction of Q is that Q is made to come as close as possible to the u-algebra of a Lie algebra, which would be a restricted extension of M by L inducing the given kernel structure on M, if it actually existed. At any rate, Q/QN will actually be the u-algebra of a restricted extension of M/N by L.

We choose a well-ordered basis (x_i) for L over F and write the elements of U_L as linear combinations of ordered monomials in these basis elements, with all exponents less than p. The elements of Q will be written in the form $\sum_k q_k u_k$, where the q_k are such ordered monomials, now regarded as elements of Q, and the u_k are arbitrary elements of the subalgebra U_M of Q. This notation will be consistent with our definition of the multiplication in Q, for we define, for $u \in U_M$, $(\sum_k q_k u_k)u = \sum_k q_k(u_k u)$. The products $u(qv)$, for u and v in U_M and q an ordered monomial, are defined inductively on the degree of q as follows: $u(x_i v) = x_i(uv) - t_{x_i}(u)v$. Suppose we have already defined $u(qv)$ whenever q is of degree no greater than some m, in such a way that $u(qv) = \sum_k q_k(u_k v)$, where each q_k is a partial product of the factors of q, and $u_k \in U_M$. Then, if r is an ordered monomial of degree $m + 1$, we have $r = x_i q$, where q is an ordered monomial of degree m. We then define $u(rv) = \sum_k x_i q_k(u_k v) - t_{x_i}(u)(qv)$. Note that, since q_k is a partial product of the factors of q, $x_i q_k$ is indeed an ordered monomial with all exponents less than p.

For each basis element x_i, we select an element x_i^* in M such that $t_{x_i}{}^p - t_{x_i}{}^{[p]} = D_{x_i^*}$, on M. Then, on U_M, $(t_{x_i}{}^p - t_{x_i}{}^{[p]})(u) = x_i^* u - u x_i^*$.

Now we define the products $x_i(qv)$ for ordered monomials q by induction on the lexicographic order $((\text{degree of } q), i)$. We may suppose that q is of degree > 0, and that the products of smaller lexicographic order than $x_i(qv)$ have already been defined. We write $q = x_j{}^e r$, where $0 < e < p$, and r is an ordered monomial containing only x_k's with $k > j$, or $r = 1$. Also, we set $s = x_j{}^{e-1} r$, so that $q = x_j s$. Now define (It is implied throughout that our definitions of products ab are extended by bilinearity to products of linear combinations of terms like a and b.) $x_i(qv) = x_i qv$, if either $i < j$, or $i = j$ and $e < p-1$; $x_i(qv) = (x_i{}^{[p]} + x_i^*)(rv)$, if $i = j$ and $e = p-1$; $x_i(qv) = x_j(x_i(sv)) + ([x_i, x_j] + \tau(x_i, x_j))(sv)$, if $i > j$.

Finally, the definition of the multiplication is completed by defining, inductively on the degree of the monomial in the first factor, $(qx_i u)(rv) = q(x_i(u(rv)))$.

It is evident from our definitions that if a and b are arbitrary elements of Q and $u \in U_M$ we have $(ab)u = a(bu)$. We shall prove also that $(au)b = a(ub)$ and $(ua)b = u(ab)$. This will be proved by induction on the degrees of the monomials occurring in a, using preliminary results which we must establish first.

If $x_i q$ is an ordered monomial beginning with x_i, and u, v are in U_M, our definitions show that $u(x_i qv) = x_i(u(qv)) - t_{x_i}(u)(qv)$. From this, it is easily shown by induction on the degree of the ordered monomials occurring in a that $u(va) = (uv)a$, for all a in Q and all u, v in U_M. Using only this and the last item of our definition of the multiplication in Q, one finds that $(au)b = a(ub)$, for all a and b in Q and all u in U_M.

Next we shall show that, for all u and v in U_M and all ordered monomials q, we have $u(x_i(qv)) = (ux_i)(qv)$. Indeed, this is clear from the above whenever $x_i q$ is an ordered monomial, i. e., whenever $q = x_j{}^e r$, where either $j > i$, or $j = i$ and $e < p-1$. Now suppose that $q = x_i{}^{p-1} r$. Then we have $u(x_i(qv)) = u(x_i{}^{[p]}(rv)) + (ux_i^*)(rv)$. Assuming our result has already been proved for monomials of smaller degree than q, this can be written

$$u(x_i(qv)) = (ux_i{}^{[p]} + ux_i^*)(rv) = (x_i{}^{[p]} u + x_i^* u - t_{x_i}{}^p(u))(rv)$$
$$= x_i(x_i{}^{p-1}(u(rv))) - t_{x_i}{}^p(u)(rv).$$

Now we have $x_i(u(rv)) = u(x_i rv) + t_{x_i}(u)(rv)$, and if we apply this $p-1$ times we obtain $x_i{}^{p-1}(u(rv)) = \sum_{k=0}^{p-1} (C_k{}^{p-1}) t_{x_i}{}^k(u)(x_i{}^{p-1-k} rv)$ $= \sum_{k=0}^{p-1} (-1)^k t_{x_i}{}^k(u)(x_i{}^{p-1-k} rv)$. If we multiply by x_i and then replace $x_i t_{x_i}{}^k(u)$

by $t_{x_i}^{k+1}(u) + t_{x_i}^{k}(u)x_i$ most of the terms from the above sum will cancel out, and we obtain $x_i(x_i^{p-1}(u(rv))) = t_{x_i}^{p}(u)(rv) + (ux_i)(qv)$. Substituting this above, we find $u(x_i(qv)) = (ux_i)(qv)$.

There remains the case $q = x_j s$, with $j < i$. Here we make the induction on the lexicographic order $((\text{degree of } q), i)$. Then, assuming the result proved in lower lexicographic order, we have $u(x_i(qv)) = u(x_j(x_i(sv)) + [x_i, x_j](sv) + \tau(x_i, x_j)(sv)) = (ux_j)(x_i(sv)) + (u[x_i, x_j] + u\tau(x_i, x_j))(sv)$. In the first term, replace ux_j by $x_j u - t_{x_j}(u)$. Then we can write it in the form $x_j(u(x_i(sv))) - t_{x_j}(u)(x_i(sv))$. By inductive hypothesis, this is equal to $x_j((ux_i)(sv)) - (t_{x_j}(u)x_i)(sv) = (x_j x_i u - x_j t_{x_i}(u) - t_{x_j}(u)x_i)(sv)$. Hence we have

$$(u(x_i(qv)) = (x_j x_i u - x_j t_{x_i}(u) - t_{x_j}(u)x_i + u[x_i, x_j] + u\tau(x_i, x_j))(sv).$$

On the other hand, using the inductive hypothesis,

$$\begin{aligned}(ux_i)(qv) &= x_i(u(qv)) - t_{x_i}(u)(qv) = x_i((ux_j)(sv)) - (t_{x_i}(u)x_j)(sv)\\ &= x_i((x_j u - t_{x_j}(u))(sv)) - (t_{x_i}(u)x_j)(sv)\\ &= x_i(x_j(u(sv))) - (x_i t_{x_j}(u) + t_{x_i}(u)x_j)(sv)\\ &= (x_j x_i u + [x_i, x_j]u + \tau(x_i, x_j)u - x_i t_{x_j}(u) - t_{x_i}(u)x_j)(sv).\end{aligned}$$

It is easy to check that the first factors in our final results for $u(x_i(qv))$ and $(ux_i)(qv)$ are equal. Hence we may conclude that $u(x_i a) = (ux_i)a$, for all a in Q and all u in U_M.

We are now in a position to prove inductively on the degree of the monomials in a that $(ua)b = u(ab)$. The critical inductive step is the following:

$$\begin{aligned}(u(x_i qv))b = ((ux_i)(qv))b &= ((x_i u - t_{x_i}(u))(qv))b\\ &= (x_i(u(qv)))b - (t_{x_i}(u)(qv))b.\end{aligned}$$

By inductive hypothesis, the second term is equal to $t_{x_i}(u)((qv)b)$. On the other hand, since $x_i q$ is an ordered monomial, it is easily seen from our definitions that the first term is equal to $x_i(u((qv)b))$, and hence, by the above, is equal to $(x_i u)((qv)b)$. Hence we have

$$\begin{aligned}(u(x_i qv))b = (x_i u - t_{x_i}(u))((qv)b) &= (ux_i)((qv)b)\\ &= u(x_i((qv)b)) = u((x_i qv)b),\end{aligned}$$

which is what was to be proved.

Now observe that our kernel structure on M induces the structure of a

restricted L-kernel in M/N. By Lemma 1.1, this restricted L-kernel M/N is extendible. Denote by (E, ϕ) the restricted extension of M/N by L which is constructed as in the proof of Lemma 1.1. The space of E is the direct sum $(L, M/N)$. We define a linear map ρ of U_L into U_E by making the image of an ordered monomial q in the basis elements of L equal to the ordered monomial in U_E which is obtained by replacing each x_i by $(x_i, 0) \in E$. Let σ denote the canonical homomorphism of U_M onto $U_{M/N} \subset U_E$. Then $\rho \otimes \sigma$ is a homomorphism of Q onto U_E whose kernel is the ideal QN of Q which is generated by the elements of N. Since U_E is associative, it follows that $a(bc) - (ab)c \in QN$, for all a, b, c in Q.

We have a linear decomposition of Q into the subspace U_L and the ideal $U_L \otimes U_{M^+}$. For any element a of Q, denote by a^0 its component in U_L, so that $a = a^0 + a^+$, where $a^+ \in U_L \otimes U_{M^+}$. We claim that there is one and only one linear map γ of QN onto N which is the identity map on N and which is such that $\gamma(ab) = a^0 \cdot \gamma(b)$, for all $b \in QN$ and all $a \in Q$, where N is regarded as a U_L-module in the natural fashion. Indeed, there can evidently be at most one such map. In order to prove its existence, it suffices to show that if (n_j) is a basis for N over F, and a_j are elements of Q such that $\sum_j a_j n_j = 0$, then each a_j must belong to $U_L \otimes U_{M^+}$. Let u_j denote the product of the finite set of elements $n_k{}^{p-1}$, where k runs over all those indices for which $a_k \neq 0$ and $k \neq j$. Since each $n_k{}^p = 0$, right multiplication by u_j gives $a_j n_j u_j = 0$, and this evidently implies that $a_j \in U_L \otimes U_{M^+}$.

We shall now have to consider multiplications in U_L and multiplications in Q simultaneously. In order to avoid confusion we shall therefore denote elements of $U_L{}^+$ by x, y, z, etc., and their corresponding elements in Q by x', y', z', etc. With this understanding, we define a trilinear function g from $U_L{}^+$ to N by $g(x, y, z) = \gamma(x'(y'z') - (x'y')z')$.

Now we observe that if any one of the elements a, b, c is in QM then $\gamma(a(bc) - (ab)c) = 0$. We use the fact that $QM = MQ = U_L \otimes U_{M^+}$, so that γ annihilates $MQN = QMN$. With $m \in M$, our assertion now follows from the relations

$$(ma)(bc) - ((ma)b)c = m(a(bc)) - (m(ab))c = m(a(bc) - (ab)c) \in MQN;$$

$$a((mb)c) - (a(mb))c = a(m(bc)) - ((am)b)c = (am)(bc) - ((am)b)c,$$

which belongs to MQN by the first case, because $am \in MQ$;

$$a(b(mc)) - (ab)(mc) = a((bm)c) - ((ab)m)c = a((bm)c) - (a(bm))c,$$

which belongs to MQN by the second case, because $bm \in MQ$.

From this we can show that g is a 3-cocycle for U_L^+ in N. In fact, we have

$$\begin{aligned}\delta g(x, y, z, t) &= x \cdot g(y, z, t) - g(xy, z, t) + g(x, yz, t) - g(x, y, zt) \\ &= \gamma(x'(y'(z't') - (y'z')t') - (xy)'(z't') + ((xy)'z')t' \\ &\quad + x'((yz)'t') - (x'(yz)')t' - x'(y'(zt)') + (x'y')(zt)').\end{aligned}$$

Since $(xy)' - x'y' \in QM$, etc., we may replace $(xy)'$ by $x'y'$, etc., in the above without changing the value of the total expression. If we make these replacements we obtain

$$\delta g(x, y, z, t) = \gamma(((x'y')z' - x'(y'z'))t').$$

Now $(x'y')z' - x'(y'z') \in QN$, and if $n \in N$ we have $nt' = t'n - t \cdot n$, whence $\gamma(ant') = 0$, for all a in Q. Hence we may conclude that $\delta g = 0$.

Next, using that $x'y' - y'x' = [x, y]' + \tau(x, y)$, for x, y in L, etc., one verifies that, for x, y, z in L,

$$\begin{aligned}g(x, y, z) - g(x, z, y) + g(z, x, y) - g(z, y, x) + g(y, z, x) - g(y, x, z) \\ = \delta\tau(x, y, z).\end{aligned}$$

This means that the primary deviation of the kernel M_g which was constructed from g in section 2 coincides with the primary deviation of M. We wish to show that M is actually equivalent to M_g, i. e., that the kernel $M + M_{-g}$ is extendible. We shall prove this by means of Lemma 1. 1. A hindrance ρ for $M + M_{-g}$ is given by $\rho(x, y) = \tau(x, y) + x \otimes y - y \otimes x$. By what we have just seen, the primary deviation $\delta\rho$ is then 0.

Put $x_i^0 = x_i^* + x_i^{p-1} \otimes x_i \in M + M_{-g}$. Then we have, in $M + M_{-g}$, $t_{x_i}{}^p - t_{x_i{}^{[p]}} = D_{x_i{}^0}$. We shall show that the secondary deviation computed from ρ and the x_i^0 is 0. It suffices to prove that $T_{x_i}(x_j) = 0$, for all i and j.

First take the case $j = i$. We have $T_{x_i}(x_i) = t_{x_i}(x_i^0) - \rho(x_i{}^{[p]}, x_i) + \sum_{k=0}^{p-1} t_{x_i}{}^k(\rho(x_i, D_{x_i}{}^{p-1-k}(x_i)))$. This reduces to $T_{x_i}(x_i) = t_{x_i}(x_i^0) - \rho(x_i{}^{[p]}, x_i)$. On the other hand, we have, in Q,

$$\begin{aligned}x_i(x_i^{p-1}x_i) - (x_ix_i^{p-1})x_i &= x_i(x_i{}^{[p]} + x_i^*) - (x_i{}^{[p]} + x_i^*)x_i \\ &= t_{x_i}(x_i^*) - \tau(x_i{}^{[p]}, x_i).\end{aligned}$$

This last expression is a secondary deviation for the kernel M, and therefore lies in N. Hence it is equal to its image under γ, i. e., to $g(x_i, x_i^{p-1}, x_i)$. On the other hand, we have, in M_{-g},

$$t_{x_i}(x_i^{p-1} \otimes x_i) = x_i{}^{[p]} \otimes x_i - x_i \otimes x_i{}^{[p]} - g(x_i, x_i^{p-1}, x_i).$$

From these two results we see at once that the above $T_{x_i}(x_i) = 0$.

The case $j \neq i$ is considerably more difficult. In order to dispose of it, we require further associativity relations in our our algebra Q. These are the following: let $a_1, \cdots, a_m$ be a set of elements of Q such that, with some given pair of indices i and j, $i \neq j$, the monomials occurring in the a's involve only x_i and x_j, the sum of their degrees in x_i being no greater than $p-1$ and the sum of their degrees in x_j being no greater than 1. Then all products $a_1 \cdots a_m$ (with all possible arrangements of the parenthesis) are equal. In order to prove this, we observe that each of these products is obtained by first writing down the factors x_i, x_j, or $u \in U_M$ in the order in which they occur in $a_1 \cdots a_m$ and then making a series of straightening operations as follows:

$$\cdots x_j x_i \cdots \rightarrow \cdots (x_i x_j + [x_j, x_i] + \tau(x_j, x_i)) \cdots, \text{ or}$$

$$\cdots u x_i \cdots \rightarrow \cdots (x_i u - t_{x_i}(u)) \cdots,$$

or according to the schemes obtained from these by interchanging i and j. Clearly, each such straightening yields a linear combination of (symbolic) products each of which is of the same type as $a_1 \cdots a_m$, the index j being replaced by some possibily different index k, resulting from $[x_j, x_i]$. Hence it will suffice to show that the order in which these straightenings are carried out is immaterial for the final result. It is clear that differences can result only from inversion of the order of the straightenings of a triple, as shown in the following scheme, where (a, b) denotes the result of straigtening the symbolic product ab, etc.:

$$\cdots abc \cdots \begin{cases} \nearrow \cdots (a,b)c \cdots \rightarrow \cdots ((a,b),c) \cdots \\ \searrow \cdots a(b,c) \cdots \rightarrow \cdots (a,(b,c)) \cdots. \end{cases}$$

If we write down all possible permutations of 3 elements from the set u, x_i, x_i, x_i, x_j, we see at once that there is only one triple in which such alternatives exist. This is ux_ix_j, if $i > j$; or ux_jx_i, if $i < j$. But, since u associates with every pair, the two straightening procedures indicated above give the same result. This completes the proof of our assertion.

What remains to be proved is that, for $j \neq i$, $T_{x_i}(x_j) = 0$. This is the sum of a secondary deviation for M and a secondary deviation for M_{-g}. We have already computed the secondary deviation for M_{-g} in terms of g. The result was $-g(x_j, x_i^{p-1}, x_i) + \sum_{k=1}^{p-1} (g(x_i^k, D_{x_i}{}^{p-1-k}(x_j), x_i) - g(x_i^k, x_i, D_{x_i}{}^{p-1-k}(x_j)))$. By the associativity relations we have just proved, the values of g which occur

in the sum are 0, except possibly for $k = p-1$. Hence the secondary deviation in M_{-g} reduces to

$$-g(x_j, x_i^{p-1}, x_i) + g(x_i^{p-1}, x_j, x_i) - g(x_i^{p-1}, x_i, x_j).$$

Now suppose first that $j > i$. Then we have, in Q, $x_i^{p-1}(x_jx_i) = (x_i^{p-1}x_j)x_i$, whence $g(x_i^{p-1}, x_j, x_i) = 0$. Similarly, $g(x_i^{p-1}, x_i, x_j) = 0$. Hence the above deviation for M_{-g} reduces to $-g(x_j, x_i^{p-1}, x_i)$. We must therefore show that the corresponding deviation for M takes the value $g(x_j, x_i^{p-1}, x_i)$. Let ζ denote commutation with x_i in Q, i. e., $\zeta(a) = x_ia - ax_i$. Consider the product $x_jx_i^{p-1}$ in Q. Since the products of at most $p-1$ factors x_i and one factor x_j are associative, we have, by an associative identity in characteristic p which we have already used in section 2, $x_jx_i^{p-1} = \sum_{k=0}^{p-1} x_i^k\zeta^{p-1-k}(x_j)$ $= x_i^{p-1}x_j + \sum_{k=0}^{p-2} x_i^k\zeta^{p-1-k}(x_j)$. Hence we obtain $(x_jx_i^{p-1})x_i = x_i^{p-1}(x_jx_i)$ $+ \sum_{k=0}^{p-2} (x_i^k\zeta^{p-1-k}(x_j))x_i$. In the first term, we may replace x_jx_i by $x_ix_j - \zeta(x_j)$; in the sum, we may re-associate the products because $\zeta^{p-1-k}(x_j)$ is of degree 1 in the x's, and then replace $\zeta^{p-1-k}(x_j)x_i$ by $x_i\zeta^{p-1-k}(x_j) - \zeta^{p-k}(x_j)$. Then we obtain $(x_jx_i^{p-1})x_i = x_i^{p-1}(x_ix_j) - \zeta^p(x_j)$. Hence we find

$$x_j(x_i^{p-1}x_i) - (x_jx_i^{p-1})x_i = t_{x_j}(x_i^*) - [x_i^{[p]}, x_j] - \tau(x_i^{[p]}, x_j) + \zeta^p(x_j).$$

If we write out $\zeta^p(x_j)$ in full, we find that the right side of the last equation is precisely the value of our secondary deviation for M. Hence the associator on the left lies in N and therefore coincides with $g(x_j, x_i^{p-1}, x_i)$. Hence the secondary deviation for M is equal to $g(x_j, x_i^{p-1}, x_i)$, as we wished to show.

There remains only the case $j < i$. In this case, $g(x_j, x_i^{p-1}, x_i) = 0$, so that the value of the secondary deviation for M_{-g} reduces to $g(x_i^{p-1}, x_j, x_i)$ $- g(x_i^{p-1}, x_i, x_j)$. Accordingly, we consider the difference of associators in Q,

$$\begin{aligned} x_i^{p-1}(x_jx_i) - (x_i^{p-1}x_j)x_i - x_i^{p-1}(x_ix_j) + (x_i^{p-1}x_i)x_j \\ = -x_i^{p-1}\zeta(x_j) - (x_i^{p-1}x_j)x_i + x_i^{[p]}x_j + x_i^*x_j. \end{aligned}$$

By an associative identity like the one we have used above, we have, with the same justification as before, $x_i^{p-1}x_j = \sum_{k=0}^{p-1} (-1)^{p-1-k}\zeta^{p-1-k}(x_j)x_i^k$, and similarly $x_i^{p-1}\zeta(x_j) = \sum_{k=0}^{p-1} (-1)^{p-1-k}\zeta^{p-k}(x_j)x_i^k$. In the first of these equations, we isolate the term with $k = p-1$ as before. Then we substitute the sums in

the above expression for the associator difference and re-associate permissibly to obtain

$$-\sum_{k=0}^{p-1}(-1)^{p-1-k}\zeta^{p-k}(x_j)x_i{}^k-\sum_{k=0}^{p-2}(-1)^{p-1-k}\zeta^{p-1-k}(x_j)x_i{}^{k+1}-(x_jx_i{}^{p-1})x_i$$
$$+x_i{}^{[p]}x_j+x_i{}^*x_j.$$

This reduces to $-\zeta^p(x_j)-x_jx_i{}^{[p]}+x_i{}^{[p]}x_j-t_{x_j}(x_i{}^*)$. As we have already seen above, this is the negative of the secondary deviation for M, and hence lies in N and so coincides with $g(x_i{}^{p-1},x_i,x_j)-g(x_i{}^{p-1},x_j,x_i)$.

Hence we may now conclude that the deviations for $M+M_{-g}$ are 0, so that $M+M_{-g}$ is extendible, i. e., M is equivalent to M_g. This completes the proof of the following main result.

THEOREM 3. 1. *The map $g\to M_g$ of the space of 3-cocycles for $U_L{}^+$ in the $U_L{}^+$-module N induces a linear isomorphism of $H_*{}^3(L,N)$ onto the space of equivalence classes of the restricted L-kernels that have N as nucleus, with $N^{[p]}=(0)$.*

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] S. Eilenberg and S. MacLane, "Cohomology theory in abstract groups," II, *Annals of Mathematics*, vol. 48 (1947), pp. 326-341.

[2] G. Hochschild, "Cohomology and representations of associative algebras," *Duke Mathematical Journal*, vol. 14 (1947), pp. 921-948.

[3] ———, "Cohomology of restricted Lie algebras," *American Journal of Mathematics*, vol. 76 (1954), pp. 555-580.

[4] M. Mori, "On the three-dimensional cohomology group of Lie algebras," *Journal of the Mathematical Society of Japan*, vol. 5 (1953), pp. 171-182.

REMARKS TO AN EARLIER NOTE (Vol. 57, pp. 539-540).*

By AUREL WINTNER.

By using the elements of the theory of normal families, the "best" domain of regularity, as determined in [2], will in this note be transferred to the case in which the differential equation is regular only in the *unknown* function. This will supply the determination of the "best" convergence interval of the expansion introduced by Perron [1].

1. Under the normalization that $(0, 0)$ is a point of the (x, y)-region on which a given function $f(x, y)$ is continuous, consider the differential equation and the initial condition

$$dy/dx = f(x, y), \qquad y(0) = 0, \tag{1}$$

and suppose that, on the real x-interval under consideration and for small $|\, y \,|$, the function $f(x, y)$ is a uniformly convergent power series in y, say

$$f(x, y) = \sum_{k=0}^{\infty} f_k(x) y^k. \tag{2}$$

Then the solution $y(x)$ of (1) is not in general a power series $a_1 x + \cdots + a_n x^n + \cdots$ but, corresponding to the mere continuity of the coefficients $f_k(x)$ of (2), a curve $y = y(x)$ about which not more than the existence of a continuous slope $dy(x)/dx$ is assured. Correspondingly, the assumptions imposed on (1) do not exclude the case, important in many applications, in which the coefficients $f_k(x)$ of (2), instead of being regular power series in x, have at $x = 0$ singularities of the type of $x^{\frac{1}{3}}$, $x \sin(x^{-1}), \cdots$ or are not analytic at all. Then such methods as that of the successive approximations (Cauchy-Lipschitz) are still applicable. But these classical methods, being applicable also when $f(x, y)$ is not a power series in y, fail to utilize the advantages which, even from the point of view of a computer of $y(x)$, might result from the power series structure of (2).

In a little-known paper, Perron [1] considered an algorithm which does not fail to utilize the assumption that $f(x, y)$ is a power series in y (but not necessarily in x). His procedure is as follows: First, replace y in (2) by ty, where t is a parameter, and, for a fixed value of t, denote by $y^t = y^t(x)$ the solution of what results from (1) upon this replacement; so that

$$dy^t/dx = f(x, ty), \qquad y^t(0) = 0. \tag{3}$$

* Received January 30, 1954.

Next, assume for the solution $y^t(x)$ of (3) a power series in t, say

$$y^t(x) = \sum_{n=0}^{\infty} \phi_n(x)\, t^n \tag{4}$$

(the coefficient functions $\phi_n(x)$ of (4) can be calculated uniquely, since substitution of (4) into (3), when followed by a comparison of the coefficients of t^n on both sides of what then becomes of (3), leads to the recursive formulae

$$\phi_0(x) = \int_0^x f_0 dx, \cdots, \phi_n(x) = \int_0^x P_n(f_0, f_1, \cdots, f_n; \phi_0, \phi_1, \cdots, \phi_{n-1})\, dx,$$

where P_n is a universal polynomial, with positive coefficients, in its $(n+1)+n$ arguments f_i, ϕ_j). Finally, observing that (3) is the case $t = 1$ of (3), hope that the case $t = 1$ of (4) will supply a convergent expansion,

$$y(x) = \sum_{n=0}^{\infty} \phi_n(x), \tag{5}$$

of the solution $y(x)$ of (1).

2. In his paper referred to, Perron justifies this hope, on the interval

$$0 \leqq x \leqq \min(a, \tfrac{1}{2}b/M), \tag{6}$$

under the following assumption: The coefficient functions $f_k(x)$ of (2) are continuous on the interval $0 \leqq x \leqq a + \epsilon$ and satisfy the inequality

$$|f_k(x)| \leqq (M - \epsilon)/b^k \text{ for } 0 \leqq x \leqq a + \epsilon, \qquad (k = 0, 1, 2, \cdots), \tag{7}$$

where a, b, M and ϵ are positive constants. In fact, Perron proves by a majorant argument that, by virtue of (7), the series (5) is absolutely convergent on the interval (6) and represents there the solution $y(x)$ of the case (2) of (1). [The letters a, b, M, used above, correspond to the standard notations, those occurring in (8) and (9) below; in these notations, Perron's four constants K, M, $b - a$, γ are represented by $M - \epsilon$, $1/b$, a, 0 respectively.]

A glance at Perron's explicit solution of his majorant of (1) shows that, under his hypothesis (7), the occurrence of the factor $\frac{1}{2}$ in the limitation (6) of the convergence range of (2) cannot be helped. The occurrence of the $\frac{1}{2}$ seems nevertheless uncalled for, since, if a, b, M have their meaning customary in the treatment of (1) by successive approximations, then the latter are known to lead to the range

$$0 \leqq x \leqq \min(a, b/M), \tag{8}$$

which does not contain the $\frac{1}{2}$ of (7) and is the "best possible" range for function-theoretical reasons; cf. [2]. It is understood that the customary meaning of the three constants a, b, M is that contained in the assumption

that (for some fixed $\epsilon > 0$ which, except for the sake of a comparison with (7), could be replaced by $\epsilon = 0$)

$$|f(x,y)| \leqq M - \epsilon \text{ for } 0 \leqq x \leqq a + \epsilon, \quad |y| < b, \tag{9}$$

where, (2) being a power series, y is allowed to be complex. Thus $|y| < b$ is a circle, and not an interval, in the second proviso of the first inequality in (9).

Accordingly, by Cauchy's estimate of the coefficients of the power series (2), Perron's inequalities (7) are implied by (9). The converse inference is false, as seen by choosing $f_k(x)$ independent of k in (2), that is, by choosing $f(x,y) = f_0(x)/(1-y)$ (if $|y| < 1$). In other words, (9) is more general than (7). This, when combined with the fact that, as emphasized in [3], Cauchy's principle of majorants can never supply "best" x-ranges of validity ("best" from the function-theoretical point of view; cf. [2]), makes one suspect that Perron's estimate (6) for a valid expansion (5) can be improved to (8), that is, to the classical estimate (8) for a valid range of successive approximations.

3. In what follows, this suspicion of a mischievous part played by Cauchy's majorants will be justified. In fact, it will be shown that if $f(x,y)$ is continuous in (x,y), satisfies the first of the inequalities (9), and is regular in y, i. e., is of the form (2), whenever x and y satisfy the conditions specified in (9) (where $y = re^{i\vartheta}$ with $r = |y| < b$, and not just $-b < y < b$, is meant), then Perron's expansion (5) is convergent on the entire interval (8) and represents there the solution $y(x)$ of (1). In addition, the convergence of (5) is absolute on (8). Incidentally, it will be clear from the proof that the convergence of (5) is uniform on (8) and that, at every point of (8), the derivative occurring in (1) can be obtained by term-by-term differentiation of (5).

Proof. The occurrence of the $\epsilon > 0$ in (9) implies the existence of a positive number λ which is somewhat greater than 1 and has the following property: If t is any complex number contained in the circle

$$|t| < \lambda, \qquad (\lambda > 1), \tag{10}$$

then

$$|f(x,ty)| < M \text{ for } 0 \leqq x < a, \quad |y| < b. \tag{11}$$

Consider (3) for a fixed t. The sequence of the standard successive approximations belonging to (3) (for a fixed t) is assigned by

$$y^t{}_m(x) = \int_0^x f(u, ty^t{}_{m-1}(u))\,du, \quad \text{where} \quad y^t{}_0(x) \equiv 0. \tag{12}$$

In view of (11), all the functions (12) are defined on the interval (8), and

$$|y^t{}_m(x)| < b \text{ if } 0 \leqq x \leqq \min(a, b/M). \tag{13}$$

Furthermore, since (2) and (11) assure the satisfaction of Lipschitz's condition,

$$y^t{}_m(x) \to y^t(x) \text{ as } m \to \infty \text{ on the interval (8)}, \tag{14}$$

where $y^t(x)$ denotes the solution of (3).

On the other hand, since $f(x, tz)$ results if y is replaced by tz in the power series (2), it is readily seen from the recursion formula (12) that each of the functions $y^t{}_0(x), y^t{}_1(x), \cdots$ is a power series in t, say

$$y^t{}_m(x) = \sum_{n=0}^{\infty} \phi^m{}_n(x) t^n, \tag{15}$$

and that, corresponding to the induction leading from (11) and (12) to (13), each of the power series (15) is convergent on the circle (10) if x is on the interval (8). Hence, for fixed x on (8), the functions $y^t{}_m(x)$ of t are regular on the circle (10). In addition, they are uniformly bounded there, by (13), and converge there to $y^t(x)$, by (14). Consequently, if x is fixed on the interval (8), then, by the elements of the theory of complex functions, $y^t(x)$ is a regular function of t on the circle (10) and possesses, therefore, an expansion (4) convergent on (10). Hence, in order to conclude the validity (and the absolute convergence) of (5) at every point x of the interval (8), it is sufficient to note that, in view of the parenthetical inequality in (10), the value $t = 1$ is allowed in (4).

It is clear from this proof that the method and the result can be generalized from (1) to the case of a system.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] O. Perron, "Ueber Integration von gewoehnlichen Differentialgleichungen durch Reihen, II," *Sitzungsberichte der Heidelberger Akademie der Wissenschaften*, vol. 1919, part A, no. 8.

[2] A. Wintner, "On the exact value of the bound for the regularity of solutions of ordinary differential equations," *American Journal of Mathematics*, vol. 57 (1935), pp. 539-540.

[3] ———, "On implicit analytic systems," *Commentarii Mathematici Helvetici*, vol. 23 (1949), pp. 294-302.

ABSTRACT PROBABILITY SPACES AND A THEOREM OF KOLMOGOROFF.* [1]

By I. E. SEGAL.

1. Introduction. In the more recent mathematical work on probability the mathematical model for a probabilistic system that is customarily employed is a "probability measure space." This consists of a set M, a σ-ring $\mathfrak{M}$ of subsets of M which includes the set M itself, and a countably additive non-negative measure m on $\mathfrak{M}$ normalized by the requirement $m(E) = 1$. In the present paper a somewhat different sort of model is presented which while equivalent to the standard one is more cogent for certain technical purposes, and seems to us to be more in agreement with the historical and conceptual development of probability theory. Among other technical advantages, the use of this model makes it possible to extend a well-known theorem of Kolmogoroff concerning the existence of random variables having preassigned joint distributions from the case of real-valued to that of general-valued random variables.

2. Abstract random variables. Random variables are usually defined as measurable functions on probability spaces. This is a long way from either the practical statistical or intuitive conceptual formulation of the notion, especially as the probability spaces required for dealing with simple concrete situations may be mathematically relatively sophisticated. For example, in treating sequences of coin tosses one is interested to begin with in the proportion of heads in n throws, in elementary algebraic expressions formed from such propositions, and in their expectations. At present however, in order to give a rigorous treatment of these simple matters, one is required to introduce the infinite direct product spaces in which each factor is a fixed two-point space, and to formulate the proportion of heads as a measurable function on this rather ungeometrical space, which must yet be endowed with the appropriate measure. A more direct approach to the matters of basic interest seems desirable, and if it is possible to avoid the explicit use of infinite product spaces, so much the better.

* Received November 10, 1953; revised March 29, 1954.

[1] This research was supported in part by The Office of Naval Research.

Such an approach can be given by starting with axioms of simple concrete significance for the random variables of interest and their expectations, rather than with relatively technical axioms concerning a measure on a space of "elementary events." Instead of deriving the random variables from the space of events, one can go in the other direction, but actually for many purposes the space of events is not only not necessary and to some extent arbitrary but introduces tiresome mathematical complications concerning sets of measure zero. To arrive at an effective set of axioms, notice that in the coin-tossing example above, the random variables that one is primarily concerned with are products of proportions of heads in n throws (for various values of n), and linear combinations of these. From the conventional "space of events" outlook, these are just the random variables that depend only on a finite number of throws. On this set $\mathfrak{R}$ of random variables there is defined a key linear functional E, the "expectation," which is determined by linearity from the knowledge of the proportion of heads in n throws and its various moments. Now whenever all such moments are given, the linear set determined by the corresponding set of random variables will be an algebra, as is in particular the case with $\mathfrak{R}$. In the general case it is reasonable to start with some undefined "random variables," all of whose moments (including joint moments) are given, and one is led immediately by forming linear combinations to an algebra with a distinguished linear functional.

Now to be specific, we define a *probability algebra* $(\mathfrak{R}, E)$ as an algebra $\mathfrak{R}$ (commutative and over the real field) on which there is a distinguished linear function E, having the following readily interpretable properties:

1) For any element a of $\mathfrak{R}$, $E(a)^2 \geqq 0$, and $E(a^2) = 0$ only if $a = 0$.

2) For any element a of $\mathfrak{R}$, there is a constant μ such that $E(ab^2) \leqq \mu E(b^2)$ for all elements b of $\mathfrak{R}$.

3) $\mathfrak{R}$ has a unit e and $E(e) = 1$.

The second axiom is the most restrictive, as it means implicitly that it is being required that the random variables in $\mathfrak{R}$ are bounded. This is actually, however, a light restriction, since in a real experiment one can measure only such random variables, and mathematically there is no difficulty in treating, as usual, the unbounded random variables in terms of the bounded ones.

The elements of $\mathfrak{R}$ are the basic random variables, there may of course be additional ones that can be obtained from the basic ones by a limiting process. Without going into detail at this point about the precise character of this limiting process, we mention that the term "probability algebra" is

justifiable along conventional lines by means of the result that the most general probability algebra is essentially an algebra of conventional random variables (i. e. measurable functions) on a probability space, and this space is essentially unique, if it be required that the random variables of the algebra determine the probability space (in e. g. the sense that the smallest σ-ring with respect to which they are all measurable is the σ-ring of the space, modulo null sets). The limit random variables obtained from the original ones then correspond to the remaining random variables on the space. The next section is devoted to the elucidation and proof of the result just quoted.

3. Structure of probability algebras. The *measure ring* of a measure space is defined as the ring of all measurable sets, modulo the ideal of all null sets, and the measure of an element of the measure ring is thus uniquely defined if set equal to the measure of any measurable set that is a representative for it. By a *random variable* on a probability space we shall mean, to be perfectly explicit, an equivalence class of real-valued measurable functions on the space, two functions being defined as equivalent if they differ only on a null set. Any collection of random variables on a probability space generates a subring of the measure ring, in the following fashion: if $f^\dagger$ is a random variable with representing measurable function f, the elements of the measure ring corresponding to sets of the form $[x: f(x) \leqq \lambda]$, for arbitrary real λ (i. e. to their residue classes modulo theideal of all null sets) constitute a σ-subring of the measure ring, which is a σ-ring. The σ-subring generated by all these subrings as $f^\dagger$ varics over the collection is the *subring of the measure ring determined by the collection.* A collection of random variables is said simply to be *determining* if the subring determined by them is the entire measure ring.

THEOREM 1. *Any probability algebra is isomorphic to a determining set of bounded random variables on a probability space, the isomorphism preserving alegbraic operations and expectation. This space is unique, in the sense that for any two such spaces there exists a probability-preserving isomorphism between their measure rings which connects their isomorphisms with the probability algebra.*

To prove the existence part of the theorem, it is convenient to introduce complex random variables, in order to make use of a known representation theorem dealing with the complex case. Let $\mathscr{C}$ be the complexification of $\mathscr{R}$, so that a general element of $\mathscr{C}$ has the form $x + iy$, with x and y in $\mathscr{R}$, and extend E to $\mathscr{C}$ by defining $E(x + iy) = E(x) + iE(y)$. It is easy to verify

that E as extended is complex linear. Now define an adjunction operation $u \to u^*$ on $\mathfrak{C}$ as follows: $(x+iy)^* = x - iy$. It is immediate that $E(u^*) = [E(u)]^*$ (where * represents complex conjugation), $E(u^*u) \geqq 0$, and $E(u^*u) = 0$ only when $u = 0$. Now introducing a function (u, v) on ordered pairs of elements in $\mathfrak{C}$ by defining $(u, v) = E(v^*u)$, it is clear that (u, v) is a positive definite inner product on $\mathfrak{C}$, with respect to which $\mathfrak{C}$ has a Hilbert space $\mathfrak{H}$ (say) as completion. The map $u \to u^*$ clearly extends to a conjugation J on $\mathfrak{H}$ and there is no difficulty in verifying that the system $(\mathfrak{H}, J, \mathfrak{C})$ is a Hilbert algebra in the sense of [4]. As it is also commutative, it follows from a well known representation theorem that there exists a measure space $\mathfrak{P} = (M, \mathfrak{M}, m)$ such that the space $L_2(\mathfrak{P})$ of square-integrable complex-valued functions is unitarily equivalent to $\mathfrak{H}$ via a unitary transformation W from $\mathfrak{H}$ onto $L_2(\mathfrak{P})$ which transforms J into complex conjugation and takes $\mathfrak{C}$ into an alegbra of bounded measurable functions on $\mathfrak{P}$ that is dense in $L_2(\mathfrak{P})$. (More precisely, we are dealing here with equivalence classes of measurable functions, but in the present proof it would seem somewhat circumlocutory to insist on the distinction.)

As an algebraic isomorphism preserves units, the unity element of $\mathfrak{C}$ maps into the unity function on M, whose square-integrability implies that $\mathfrak{P}$ is finite. More precisely, for any element u of $\mathfrak{C}$, $E(u) = (u, e) = (Wu, We)$ and substituting $u = e$ it results that $\| We \|^2 = 1$, i. e. the integral over M of the unity function is one, so that $\mathfrak{P}$ is a probability space. Now the image of $\mathfrak{C}$ is dense in $L_2(\mathfrak{P})$, and the desired result now follows from the fact that any dense subset of $L_2(\mathfrak{P})$ is determining. To see this, note that the functions in $L_2(\mathfrak{P})$ that are measurable with respect to the ring of sets determined by a dense set $\mathfrak{D}$ constitute a closed linear manifold in $L_2(\mathfrak{P})$, by the Riesz-Fischer theorem. This manifold is a proper submanifold if $\mathfrak{D}$ is not determining, but it cannot be proper since it contains $\mathfrak{D}$, which is dense in $L_2(\mathfrak{P})$.

To prove uniqueness, suppose that $\mathfrak{P}' = (M', \mathfrak{M}', m')$ is a space with the same properties as $\mathfrak{P}$. Let $\mathfrak{D}$ and $\mathfrak{D}'$ be the respective subsets of $L_2(\mathfrak{P})$ and $L_2(\mathfrak{P}')$ which map onto $\mathfrak{C}$ via the respective isomorphisms W and W'. The $W^{-1}W'$ is an algebraic isomorphism V of $\mathfrak{D}$ onto $\mathfrak{D}'$. It suffices to show that V extends to a unitary map of $L_2(\mathfrak{P})$ onto $L_2(\mathfrak{P}')$ that is multiplicative on bounded functions. For such a map takes idempotents into idempotents, and so will take a characteristic function C_A into a characteristic function $C_{A'}$. As the map is unitary, A and A' will have the same probability, and the induced map $A \to A'$ is unique modulo null sets. It is thus a probability-preserving map of the measure ring of $\mathfrak{P}$ into that of $\mathfrak{P}'$, and it follows from

the fact that V is an algebraic isomorphism on the bounded functions that this induced map is a ring isomorphism of the measure ring of $\mathcal{P}$ onto that of $\mathcal{P}'$. It is well known that such a map has a unique extension to an algebraic isomorphism of the random variables on $\mathcal{P}$ onto the random variables on $\mathcal{P}'$.

Now to show that V has an extension of the sort indicated, note first that it has a unique unitary extension V_1 from $L_2(\mathcal{P})$ onto $L_2(\mathcal{P}')$, as it is isometric and maps a dense subset of $L_2(\mathcal{P})$ onto a dense subset of $L_2(\mathcal{P}')$. Next, if f is a bounded function on $\mathcal{P}$ and g is in $\mathcal{E}$, then $V_1(fg) = V_1(f)V_1(g)$. For let $\{f_n\}$ be a sequence of elements of $\mathcal{E}$ which converges in $L_2(\mathcal{P})$ to f. Then as g is bounded, $\{f_n g\}$ converges to fg in $L_2(\mathcal{P})$, so that $V_1(f_n g)$ converges to $V_1(fg)$ in $L_2(\mathcal{P}')$. On the other hand, $V_1(f_n g) = V_1(f_n)V_1(g)$, and as $V_1(g)$ is bounded and $\{V_1(f_n)\}$ converges to $V_1(f)$ in $L_2(\mathcal{P}')$, it results that $V_1(f_n g)$ converges to $V_1(f)V_1(g)$. The stated equality now follows directly. Now the set of all elements g of $L_2(\mathcal{P})$ for which $V_1(fg) = V_1(f)V_1(g)$ for every bounded function f on $\mathcal{P}$ is a closed subset of $L_2(\mathcal{P})$, since either side of this equality is a continuous function of g. As it includes the dense subset $\mathcal{E}$, it must be all of $L_2(\mathcal{P})$.

Remark 1. It was noted above that a dense subset of $L_2(\mathcal{P})$ is determining. Conversely, *a determining subset of $L_2(\mathcal{P})$ is dense, if it is an algebra of bounded functions.*[2] To establish this measure-theoretic analog to the Stone Weierstrass theorem, let S denote the collection of all finite unions of sets of the form $[x: f_i(x) \in A_i \ (i = 1, 2, \cdots, n)]$, where A_i is an interval (of any type) and the f_i are elements of $\mathcal{D}$. It is clear that $\mathcal{S}$ is closed under unions, and it is closed under differences, since

$$\bigcap_i [x: f_i(x) \in A_i] - \bigcap_j [x: g_j(x) \in B_j] = \bigcap_{i,j} [x: f_i(x) \in A_i] \cap [x: g_j(x) \in' B_j],$$

and the complement of B_j is the union of at most two intervals. Thus $\mathcal{S}$ is

[2] This result is stated by Kakutani in *Proceedings of the National Academy of Sciences*, vol. 36 (1950), p. 320, without the restriction to bounded functions. That it is invalid in this generality can be deduced from the non-uniqueness of the Stieltjes moment problem in general. The latter implies the existence of a function F of bounded variation on $(0, \infty)$ such that $\int_0^\infty \lambda^n dF(\lambda) = 0 \ (n = 1, 2, \cdots)$, while $dF(\lambda) \not\equiv 0$. Putting $G(\mu) = \int_0^\mu |dF(\lambda)|$, the algebra of polynomials on $(0, \infty)$ vanishing at 0 determines the ring of Borel sets on $(0, \infty)$ but cannot be dense in the space of functions square-integrable with respect to the measure determined by G, as every element is orthogonal to dF/dG.

a Boolean ring. It is clear that the σ-ring $\mathfrak{R}$ generated by $\mathfrak{S}$ is the same as the σ-ring determined by $\mathfrak{D}$. Now if A is a finite interval, the characteristic function of A is a uniformly bounded limit of polynomials, which implies that the characteristic function of $[x: f(x) \in A]$ is a uniformly bounded limit of a sequence of elements of $\mathfrak{D}$, if f is in $\mathfrak{D}$.

Now if g is an element of $L_2(\mathfrak{P})$ that is orthogonal to all the elements of $\mathfrak{D}$, it is orthogonal to all uniformly bounded limits of sequences of elements of $\mathfrak{D}$, and so in particular, $\int_Q g = 0$, if $Q = [x: f_i(x) \in A_i \ (i = 1, 2, \cdots, n)]$, where the f_i are in $\mathfrak{D}$ and the A_i are bounded intervals. By the countable additivity of the indefinite integral, the same holds when the A_i are not necessarily bounded. It follows that $\int_S g = 0$ for all elements S of $\mathfrak{S}$. As the collection of all measurable sets S such that $\int_S g = 0$ is closed under unions of monotone sequences and contains the ring $\mathfrak{S}$, it follows that this collection includes the σ-ring generated by $\mathfrak{S}$. Now every measurable set is equivalent to some element of $\mathfrak{R}$, since $\mathfrak{D}$ is determining, so that $\int_S g = 0$ for every measurable set S. This means that $g = 0$, i. e. $\mathfrak{D}$ must be dense.

In a similar fashion it can be shown that a *determining algebra of bounded measurable function on P is dense in the algebra of all bounded measurable functions* of $\mathfrak{P}$, in the topology on this algebra induced from the weak topology on the dual of the Banach space $L_1(\mathfrak{P})$, by the identification of a linear functional with the bounded measurable function that is its kernel in the Riesz representation for the functional.

Remark 2. An isomorphism of the type involved in the uniqueness part of Theorem 1 automatically has further properties justifying the identification of the two spaces from a probability viewpoint. It can be uniquely extended to an algebraic isomorphism between the random variables on the respective spaces, that preserves order (non-negatively) and expectation (i. e. maps a random variable of finite expectation into a random variable of the same finite expectation).

4. Existence of random variables with preassigned distributions. A well known theorem of Kolmogoroff asserts that there exist real-valued random variables having any preassigned joint distributions satisfying the obviously necessary consistency conditions. The proof of the theorem does not extend

to the case of general-valued random variables, the basic difficulty being that a measure that is countably additive on each of a monotone increasing sequence of σ-rings of subsets of a set need not be countably additive on their union. This difficulty essentially disappears when the theorem is approached from the standpoint of probability algebras. This is because the trouble in the conventional proof arises, roughly speaking, from trying to hang the random variables on the wrong space, while the use of probability algebras leaves the choice of the space of events quite arbitrary, within wide limits.

In order to deal effectively with general-valued random variables, a modification of the usual notion must be introduced. The basic conceptual concern in statistics is not so much with the values of the measurable function f representing a random variable at points of the space M of events as with the sets in M where f takes on certain values (and with the probabilities of those sets). That is, the inverse map $F = f^{-1}$ from (Borel) sets on the reals to subsets of M is more central from some viewpoints associated with the origin of the theory. Taking the values of F modulo null sets, one obtains a Boolean σ-homomorphism $F^\dagger$ from the Borel subsets of the reals to the measure ring of $\mathfrak{P}$, which determines uniquely the random variable $f^\dagger$ represented by f, and every such σ-homomorphism preserving the Boolean units arises in this way. Thus $f^\dagger$ and $F^\dagger$ are mathematically equivalent notions, and there are some mathematical as well as conceptual advantages in using $F^\dagger$, particularly in the case of general-valued random variables. Accordingly we take

Definition. *A generalized random variable* X on a probability space $\mathfrak{P}$ is a Boolean-σ-homomorphism from a Boolean σ-ring $\mathfrak{B}$ with unit to the measure ring of $\mathfrak{P}$, which preserves the Boolean units. If $\mathfrak{B}$ consists of subsets of a set B, then X is called B-valued, and in any case is said to be *defined relative to* $\mathfrak{B}$.

The usual notion of joint probability distribution of a set of random variables extends to generalized random variables as follows. For generalized random variables X_i defined relative to $\mathfrak{B}_i$ $(i = 1, 2, \cdots, n)$, the *joint probability* that X_i has its value in b_i $(i = 1, 2, \cdots, n;\ b_i \in \mathfrak{B}_i)$ is defined as the probability of the element $\bigcap_{i=1}^{n} X_i(b_i)$ of the measure ring of the probability space. This probability, as a function of $b_1, b_2, \cdots, b_n$, is the *joint distribution* of $X_1, X_2, \cdots, X_n$.

Kolmogoroff's theorem may now be extended as follows.

THEOREM 2. *Let $\mathfrak{B}_\lambda$ be a Boolean σ-ring with unit u_λ, where λ ranges over an index set Λ. For each finite subset $\lambda_1, \lambda_2, \cdots, \lambda_n$ of Λ let there be given a function $m_{\lambda_1,\lambda_2,\cdots,\lambda_n}$ on the (set-theoretic) direct product $\prod_{k=1}^{n} \mathfrak{B}_{\lambda_k}$ to the interval $[0, 1]$. In order that there exist a probability space $\mathfrak{P}$ and generalized random variables X_λ on $\mathfrak{P}$ defined relative to $\mathfrak{B}_\lambda$ $(\lambda \in \Lambda)$ such that the joint distribution of any finite set of the $X_{\lambda_1}, X_{\lambda_2}, \cdots, X_{\lambda_n}$ be $m_{\lambda_1,\lambda_2,\cdots,\lambda_n}$ it is necessary and sufficient that:*

1) $m_{\lambda_1,\lambda_2,\cdots,\lambda_n}(u_{\lambda_1}, u_{\lambda_2}, \cdots, u_{\lambda_n}) = 1$.

2) *if the elements b_{λ_j} of B_{λ_j} $(j = 1, 2, \cdots, n)$ are fixed for $j \neq k$, then $m_{\lambda_1,\lambda_2,\cdots,\lambda_n}(b_{\lambda_1}, b_{\lambda_2}, \cdots, b_{\lambda_n})$ is countably additive as a function of b_{λ_k}.*

3) *if $b_{\lambda_j} = u_{\lambda_j}$ for $j > k$, then*

$$m_{\lambda_1,\lambda_2,\cdots,\lambda_n}(b_{\lambda_1}, b_{\lambda_2}, \cdots, b_{\lambda_n}) = m_{\lambda_1,\lambda_2,\cdots,\lambda_k}(b_{\lambda_1}, b_{\lambda_2}, \cdots, b_{\lambda_k}).$$

The necessity of these conditions is obvious, as in the case of real-valued random variables. In proving the sufficiency we use

LEMMA 2.1. *The simple functions modulo null functions on a finitely-additive finite normalized measure space constitute a probability algebra.*

The distinguished linear functional of the probability algebra in the lemma is understood to be the integral. If $Q = (G, \mathfrak{G}, g)$ is a finitely-additive finite normalized measure space—i. e. g is a finitely-additive non-negative measure on the Boolean ring $\mathfrak{G}$ of subsets of G, where G is in $\mathfrak{G}$ and $g(G) = 1$—a simple function is a finite linear combination of characteristic functions of sets in $\mathfrak{G}$. Such a function is a null function if it vanishes except on a set on which g vanishes. The simple functions constitute a ring in which the null functions are an ideal, and the quotient algebra $\mathfrak{A}$ is the algebra of the probability algebra whose existence is asserted by the lemma.

If the simple function f has the form $f = \sum_{i=1}^{n} \alpha_i c_{A_i}$ with A_i in $\mathfrak{G}$, the integral $E_0(f)$ is defined as $\sum_{i=1}^{n} \alpha_i g(A_i)$, and there is no difficulty in verifying that E_0 is linear on the simple functions, that $E_0(f^2) \geqq 0$ while $E_0(f^2) = 0$ only if f is a null function, and that $E_0(f^2 h) \leqq \mu E_0(f^2)$ for arbitrary simple f and fixed simple h, μ being a constant depending on h. If $f^\dagger$ denotes the equivalence class represented by f, and if E is defined on $\mathfrak{A}$ by setting $E(f^\dagger) = E_0(f)$, it follows that E is uniquely defined and that $(\mathfrak{A}, E)$ is a probability algebra.

Proof of Theorem 2. It is well known that a Boolean σ-ring $\mathfrak{B}_\lambda$ is isomorphic to a σ-ring $\mathfrak{C}_\lambda$ of subsets of a set B_λ, modulo a σ-ideal $\mathfrak{D}_\lambda$ in it, where B_λ is a representative for the unit of $\mathfrak{C}_\lambda/\mathfrak{D}_\lambda$. It is therefore no loss of generality to assume that $\mathfrak{B}_\lambda = \mathfrak{C}_\lambda/\mathfrak{D}_\lambda$. Now set $B = \prod_{\lambda \in \Lambda} B_\lambda$, and let $\mathfrak{R}$ be the ring of all finite unions of "cylinder sets," i. e. sets of the form $\prod_{\lambda \in \Lambda} A_\lambda$, where $A_\lambda \in \mathfrak{C}_\lambda$ for all λ and $A_\lambda = B_\lambda$ except for finitely many λ. Define a functional m on cylinder sets by setting

$$m(\prod_{\lambda \in \Lambda} A_\lambda) = m_{\lambda_1, \lambda_2, \cdots, \lambda_n}(A_1, A_2, \cdots, A_{\lambda_n})$$

where $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ are the indices for which $A_\lambda \neq B_\lambda$, and setting $m(B) = 1$ to cover the case where there are no such indices. Then m has a unique finitely additive extension to the ring $\mathfrak{R}$, as is well known. Applying the preceding lemma and Theorem 1, it results that there exists a probability space $\mathfrak{P}$ and a determining algebra $\mathfrak{A}'$ of random variables on $\mathfrak{P}$ that is isomorphic (both as regards algebraic operations and the respective distinguished linear functionals) to the algebra of simple functions modulo null functions on $(B, \mathfrak{R}, m)$.

For each fixed μ in the index set Λ, a generalized random variable X_μ, on $\mathfrak{P}$ and relative to $\mathfrak{B}_\mu$, may now be defined as follows. Let a_μ be an element of $\mathfrak{B}_\mu$, let A_μ be a representing subset of B_μ for a_μ, put f_μ for the equivalence class of the characteristic function of $A'_\mu = \prod C_\lambda$, where $C_\lambda = B_\lambda$ for $\lambda \neq \mu$ and $C_\mu = A_\mu$, and let f'_μ denote the image in $\mathfrak{A}'$ of f_μ under the isomorphism just described. As f_μ is idempotent, so also is f'_μ which implies that there exists a unique element a'_μ of the measure ring of $\mathfrak{P}$ such that f'_μ is the equivalence class of the characteristic function of any representing set for a'_μ. The map $a_\mu \to a'_\mu$ is designated as X_μ. From the fact that the preceding isomorphism is in particular an isomorphism when restricted to characteristic functions modulo null functions, it follows directly that X_μ is a Boolean homomorphism of $\mathfrak{B}_\mu$ into the measure ring of $\mathfrak{P}$. It is clear that it preserves the Boolean units involved. To show that it is actually a σ-homomorphism, let $a_\mu^{(1)}, a_\mu^{(2)}, \cdots$ be a sequence of disjoint elements of $\mathfrak{B}_\mu$ of union a_μ. Then $a_\mu \geqq \bigcup_{i=1}^{n} a_\mu^{(i)}$, so that $X_\mu(a_\mu) \geqq \bigcup_{i=1}^{n} X_\mu(a_\mu^{(i)}$, since X_μ is a homomorphism, and letting n become infinite, it follows that $X_\mu(a_\mu) \geqq \bigcup_{i=1}^{\infty} X_\mu(a_\mu^{(i)})$. Now if two comparable elements of the measure ring of $\mathfrak{P}$ have the same measure, they must be equal, so it suffices now to show that $m'(X_\mu(a_\mu)) = m'(\bigcup_{i=1}^{\infty} X_\mu(a_\mu^{(i)}))$, where m' denotes the measure function on

the measure ring of $\mathfrak{P}$. By construction, $m'(X_\mu(a_\mu)) = m_\mu(a_\mu)$, while $m'(\bigcup_{i=1}^{\infty} X_\mu(a_\mu^{(i)})) = \sum_{i=1}^{\infty} m'(X_\mu(a_\mu^{(i)}))$ (for m' is countably additive and the $X_\mu(a_\mu^{(i)})$ are mutually disjoint, X_μ being a homomorphism) $= \sum_{i=1}^{\infty} m_\mu(a_\mu^{(i)})$ $= m_\mu(a_\mu)$ by the countable additivity of m_μ.

It remains only to show that

$$m'(\bigcap_{j=1}^{n} X_{\lambda_j}(a_{\lambda_j})) = m_{\lambda_1,\lambda_2,\cdots,\lambda_n}(a_{\lambda_1}, a_{\lambda_2}, \cdots, a_{\lambda_n})$$

for arbitrary $a_{\lambda_1}, a_{\lambda_2}, \cdots, a_{\lambda_n}$ in $\mathfrak{B}_{\lambda_1}, \mathfrak{B}_{\lambda_2}, \cdots, \mathfrak{B}_{\lambda_n}$ respectively. Using again the fact that the isomorphism involved above is in particular a ring isomorphism on the equivalence classes of the characteristic functions, and the definition of the X_λ, the equivalence class of the characteristic function of $\bigcap_{j=1}^{n} X_{\lambda_j}(a_{\lambda_j})$ corresponds to the equivalence class of the characteristic functions of $\bigcap_{j=1}^{n} A'_{\lambda_j}$. As the isomorphism preserves the respective distinguished linear functionals,

$$m'(\bigcap_{j=1}^{n} X_{\lambda_j}(a_j)) = m(\bigcap_{j=1}^{n} A'_{\lambda_j}) = m_{\lambda_1,\lambda_2,\cdots,\lambda_n}(a_{\lambda_1}, a_{\lambda_2}, \cdots, a_{\lambda_n}).$$

5. Weak distributions on linear spaces. In the case of linear-space valued random variables a special approach may be appropriate. In certain situations, the relevant object is not actually a probability distribution on the linear space, but a kind of weak distribution. For example, it is well known that there exist no probability distributions on Hilbert space having incisive invariance and smoothness properties. Nevertheless there exists a weak distribution on Hilbert space which is significant for applications to quantum field theory and to the theory of Brownian motion.

Definition. A *weak distribution* on a topological linear space B is a linear map $u^* \to F(u^*)$ from the space B^* of continuous linear functionals on B to the random variables on a probability space.

In case B is finite-dimensional, there is clearly a unique distribution of the conventional type on B, such that the identity map on B^* coincides with the given weak distribution, but in the infinite-dimensional case this need not be the case. In the case of a (real) Hulbert space $\mathfrak{H}$, there is an essentially unique weak distribution on $\mathfrak{H}$ such that (identifying $\mathfrak{H}^*$ with $\mathfrak{H}$ in the usual way) $F(x)$ is normally distributed with norm zero and variance $\| x \|^2$, while $F(x_1), \cdots, F(x_n)$ are independently distributed whenever $x_1, \cdots, x_n$ are mutually orthogonal. When $\mathfrak{H}$ is finite-dimensional, this is

essentially (i. e. via the correspondence between weak and ordinary distributions just mentioned) the normal multivariate distribution in the standard form, while for infinite-dimensional $\mathfrak{H}$ it constitutes a proper generalization of the normal distribution.

The question of the existence of weak distributions with preassigned joint distribution for $F(u_1^*), \cdots, F(u_n^*)$, $u_1^*, \cdots, u_n^*$ being an arbitrary finite set of elements in B^*, can be resolved by the use of the original theorem of Kolmogoroff: *it is necessary and sufficient for such existence that the usual consistency conditions hold, and moreover that if* $w^* = \alpha u^* + \beta v^*$, *then the prescribed joint distribution of* $F(u^*)$, $F(v^*)$ *and* $F(u^*)$ *is such that* $F(w^*) = \alpha F(u^*) + \beta F(v^*)$. To derive the existence of the normal weak distribution on Hilbert space described above, the joint distribution of $F(u_1), \cdots, F(u_n)$ is prescribed as normal, with norm zero, and with covariance matrix $\| (u_i, u_j) \|$. The validity of the general consistency conditions follows at once from well known properties of the normal distribution, as does also the special condition for linearity.

6. Concluding discussion. The present somewhat abstract point of view regarding variables has further theoretical advantages, in such areas as measurability problems in the theory of continuous stochastic processes, and in reformulating the theory in such a way that it can be extended to the non-commutative case.

In these measurability problems the key difficulty is that of obtaining a unique and measurable least upper bound for a non-denumerable collection of random variables. If the random variables are formulated as measurable functions, the pointwise least upper bound will in general not be measurable, and in any case will depend upon the specific representations for the random variables that are selected. On the other hand, due to the completeness of abstract random variables as a partially ordered set, the least upper bound of an arbitrary collection of abstract random variables always exists uniquely as an abstract random variable. This approach has been applied by Nelson [3] to give a more intrinsic approach to such questions as the continuity of sample functions. A similar method could be used to treat generalizations of such questions as continuity of the sample functions in the case of generalized random variables, although in this case there is no analog for the sample functions; this is possible because there are analogs for the statistically significant aspects of the sample functions.

The non-commutative theory essentially arises if the assumption of commutativity is omitted in the axioms for a probability algebra, and if

corresponding assumptions made in the assumptions on the distinguished linear functional. This transition from a commutative algebra to a non-commutative one is precisely the transition from a conventional probability space to a quantum-mechanical system in a definite state. In the one case it develops that the elements of the algebra can be represented by measurable functions on a measure space, while in the other it develops that they can be represented by operators on a Hilbert space. It is convenient in the non-commutative case to deal with a complex algebra admitting an adjunction operation *, although the conceptual loss thereby is to a considerable extent avoidable. In this case it is appropriate to assume, as is the case in the conventional scheme of quantum mechanics, that the distinguished linear functional E has the properties: $E(aa^*) \geqq 0$, $E(b^*ab) \leqq \mu_a E(b^*b)$, and $E(e) = 1$. The basic notions of probability theory, including conditional expectation can be extended to such a system, which always admits a representation in which the elements of the algebra are operators on a Hilbert space and E has the form $E(A) = (A\psi, \psi)$ for a fixed unit vector ψ. The theory remains fairly rudimentary, however, unless it is assumed further that $E(ab) = E(ba)$, in which case most probabilistic concepts and many results can, with appropriate reformulation be extended. Among the interesting examples of such systems are the approximately finite factors of Murray and von Neumann; these are essentially direct limits of matrix algebras, the distinguished linear functional being the corresponding limit of the traces on the matrix algebra, after normalization so that the unit matrix has unit trace.

COLUMBIA UNIVERSITY, NATIONAL UNIVERSITY OF MEXICO, AND
UNIVERSITY OF CHICAGO.

REFERENCES.

[1] J. L. Doob, *Stochastic Processes*, New York, 1953.

[2] A. Kolmogorov, *Grundbegriffe der Wahrscheinlichskeitsrechnung*, Ergebnisse der Mathematik, vol. 2 (1933).

[3] E. Nelson, "An invariant condition for sample function continuity," to appear.

[4] I. E. Segal, "A non-commutative extension of abstract integration," *Annals of Mathematics*, vol. 57 (1953), p. 401.

OHIO UNIVERSITY
LIBRARY

THEORY OF POSITIVE LINEAR DEPENDENCE.*

By CHANDLER DAVIS.

1. Introduction. In problems involving real linear spaces, the concept of a positive linear combination of a set of vectors has acquired an importance almost comparable to that of an (unrestricted) linear combination. As examples, one notes the usefulness of the Kreĭn-Milman-Šmulyan theorem in functional analysis; and the recent theories of games and linear programming. Accordingly, it seems desirable to develop an analague to the standard theory of linear subspaces and linearly independent sets of vectors. The present paper is an attempt in this direction.

A central concern, accordingly, will be *the geometry of frames of solid cones*[1] in finite-dimensional Euclidean space (they being the analogues of bases in the standard theory). This is a problem of wider interest than might appear in connection with frames of arbitrary convex polyhedral cones. Indeed, it is known (and easy to show) that a frame of any such cone consists of a frame of its lineality space (therefore, a frame of a solid cone of lower dimension), together with a set of vectors arbitrary except for the requirement that it map onto a frame of the cone's image when the space is reduced modulo the lineality space. Therefore, there are really only the two extreme cases to study: frames of pointed cones, and frames of solid cones. The first problem reduces in turn to the study of arbitrary convex polyhedra; the second is attacked here.

After certain necessary preliminaries (§ 2) I continue in § 3 with those parts of the theory which can be easily developed in rough analogy with classical methods in the standard (unrestricted-linear) theory. These results are all essentially known; many of them appear in [7]. The structure of bases in the present case turns out to be quite complicated, but by means of an auxiliary construction (§ 5) they are classified. Some numerical examples are collected in § 7.

* Received May 18, 1953; revised January 4, 1954.

[1] Terminology in this paragraph follows [6]. "Frame" is defined in § 2. The *lineality space* of a cone is the maximal linear subspace it contains; the lineality space is zero for a *pointed* cone, while for a *solid* cone it is the whole space.

2. Preliminaries. The space studied is real linear space of n dimensions, which will be thought of as Euclidean space E^n. Its elements will be denoted by lower-case letters; real numbers by lower-case Greek letters. The inner product of a and b will be written $a \cdot b$.

A *positive combination* of $a_1, \cdots, a_r$ is a linear combination $\lambda_1 a_1 + \cdots + \lambda_r a_r$ with $\lambda_i \geqq 0$; if all $\lambda_i > 0$ and all $a_i \neq 0$, we speak of a *strictly positive combination.* An equation giving 0 as a positive combination of certain elements is a *positive relation* between those elements; *strictly positive relations* are defined similarly.

A convex polyhedral cone, or simply a *convex cone,* is the set of all positive combinations of a finite set of vectors, not all zero. This set of vectors will be said to positively span the convex cone; or simply to *span* it. (No confusion will arise with the notion "linearly span.")

A set of vectors $\{a_1, \cdots, a_r\}$ is *positively dependent* if one of the a_i is a positive combination of the others (hence, in particular, if any a_i is zero). Otherwise the set is *positively independent.*

$\{a_1, \cdots, a_r\}$ is a *frame* of a cone C provided it is positively independent and spans C. If C is the whole space, or a linear subspace, any frame of C will be called a positive basis, or simply a *basis.*

Other non-standard terminology will be explained as it is introduced.

In selecting, from a set $\{a_1, \cdots, a_r\}$, an a_i having a specified property, I will sometimes assume without apology (or loss of generality) that the one selected is a_1. (See for example Theorem 3. 7.)

3. Some simple theorems.

THEOREM 3. 1. *$\{a_1, \cdots, a_r\}$ spans E^n if and only if, for every non-zero b, there exists an $i = 1, \cdots, r$ such that $b \cdot a_i > 0$.*

Proof. "Only if": Any b may be expressed as $b = \lambda_1 a_1 + \cdots + \lambda_r a_r$, with $\lambda_i \geqq 0$. Now $0 < b \cdot b = \lambda_1 a_1 \cdot b + \cdots + \lambda_r a_r \cdot b$, so some term in the sum must be positive.

"If": Assume the convex cone A spanned by $\{a_1, \cdots, a_r\}$ is not E^n. Then (see [6] or [4]), there is some hyperplane such that A lies entirely on one side of it; that is, there is some $b \neq 0$ such that $a \in A$ implies $b \cdot a \leqq 0$.

THEOREM 3. 2. *The set $\{a_1, \cdots, a_r\}$ is positively independent if and only if no proper subset spans the same convex cone.*

The proof is trivial and may be omitted.

THEOREM 3.3. *a_1 is a positive combination of $\{a_2, \cdots, a_r\}$ if and only if, for all b such that $b \cdot a_1 > 0$, there is an $i = 2, \cdots, r$ such that $b \cdot a_i > 0$. $\{a_1, \cdots, a_r\}$ is positively independent if and only if, for every $i = 1, \cdots, r$, there is some b such that $b \cdot a_i > 0$, but, for $j \neq i$, $b \cdot a_j \leqq 0$.*

Using the Minkowski-Weyl Theorem ([6], Theorem 11), it is not hard to reduce this to Theorem 3.2.

THEOREM 3.4. *Any subset of a positively independent set of vectors is positively independent.*

THEOREM 3.5. *Any set having a subset which spans E^n, itself spans E^n.*

These are obvious.

THEOREM 3.6. *Let $\{a_1, \cdots, a_r\}$ with $a_i \neq 0$, linearly span E^n. Then the following are equivalent:*

(i) *$\{a_1, \cdots, a_r\}$ positively spans E^n;*

(ii) *for every $i = 1, \cdots, r$, $-a_i$ is in the convex cone spanned by the remaining a_j;*

(iii) *there is some strictly positive relation* (see § 2) *between the a_i, $i = 1, \cdots, r$.*

Proof. Assume (i). Certainly $-a_1$ is expressible as $-a_1 = \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_r a_r$, with $\lambda_i \geqq 0$; if this equation is solved for a_1 and the sign changed, it will be seen that $-a_1$ is given as a positive combination of $a_2, \cdots, a_r$. (ii) is proved.

Assume (ii). It tells us there are relations

$$(1 + \lambda_{11})a_1 + \lambda_{12}a_2 + \cdots + \lambda_{1r}a_r = 0,$$

$$\lambda_{21}a_1 + (1 + \lambda_{22})a_2 + \cdots + \lambda_{2r}a_r = 0,$$

$$\lambda_{r1}a_1 + \lambda_{r2}a_2 + \cdots + (1 + \lambda_{rr})a_r = 0,$$

with $\lambda_{ij} \geqq 0$. Adding these equations gives the required expression

$$\mu_1 a_1 + \cdots + \mu_r a_r = 0, \qquad \mu_i > 0, \tag{1}$$

and (iii) is proved.

Finally, assume (iii), and let b be an arbitrary vector of E^n. Since $\{a_1, \cdots, a_r\}$ linearly spans E^n (this hypothesis enters here for the first time), there is some relation $\nu_1 a_1 + \cdots + \nu_r a_r = b$. By adding to this equation a sufficiently high positive multiple of (1), we get an expression of b as a

positive combination of $\{a_1, \cdots, a_r\}$. (i) is established and the proof is complete.

THEOREM 3.7. *If $\{a_1, \cdots, a_r\}$ positively spans E^n, then $\{a_2, \cdots, a_r\}$ linearly spans E^n.*

Proof omitted.

In the remainder of this paper, $\{a_1, \cdots, a_r\}$ will always be assumed to be a basis; when other sets of vectors are discussed different symbols will be chosen.

It is clear that the standard theorem that any two linear bases of E^n are affine equivalent has no analogue for positive bases. Thus in E^3 there are four essentially different types of positive basis, represented in Examples I-IV. (It is not true, of course, that any positive basis of E^3 is affine equivalent to one of Examples I-IV. These four examples are essentially given in [8].) These bases do not agree even in the number of vectors, which ranges from 4 to 6. Also, it is evidently not true that any subset of a positive basis is a positive basis of the smallest linear subspace containing it. The situation will become more complicated than in the case of linear dependence.

However, one sees immediately the following

THEOREM 3.8. *$r \geqq n+1$, and $r = n+1$ is possible.*

Proof. If $r \leqq n$, then either $\{a_1, \cdots, a_r\}$ fails to span E^n linearly, or it is exactly a linear basis of E^n. In either case it fails to span E^n positively. On the other hand, for any n one can choose $\{a_1, \cdots, a_n\}$ a linear basis and $a_{n+1} = -\lambda_1 a_1 - \cdots - \lambda_n a_n$, $\lambda_i > 0$, making $\{a_1, \cdots, a_{n+1}\}$ a positive basis.

A basis for which $r = n+1$ will be called *minimal*. By Theorems 3.6 and 3.7, every minimal basis is of the type just constructed.

4. Spanned subspaces. Let us therefore investigate non-minimal bases.

THEOREM 4.1. *The following are equivalent:*

(i) $r > n+1$;

(ii) *some proper subset of $\{a_1, \cdots, a_r\}$ is linearly dependent;*

(iii) *some proper subset of $\{a_1, \cdots, a_r\}$ is a positive basis of a linear subspace* (which by Theorem 3.2 must be proper) *of E^n.*

Furthermore, the subset in (iii) *may be chosen to include any preassigned a_i, and the basis may be required to be minimal.*

Proof. (iii) immediately implies (ii); and one concludes easily from Theorem 3.7 that (ii) implies (i).

Assume $r > n+1$. The vectors $a_2, \cdots, a_r$ are more than n in number, so there is some non-trivial linear relation between them. By adding the proper multiple of this equation to the equation (1) which must hold, one can certainly arrange to obtain $\mu_1' a_1 + \cdots + \mu_r' a_r = 0$, with $\mu_1' = \mu_1$, $\mu_i' \geqq 0$, but some μ_i' zero, say $\mu_{p+1}' = \mu_{p+2}' = \cdots = \mu_r' = 0$.

$\{a_1, \cdots, a_p\}$ linearly spans some subspace S of E^n. The relation $\mu_1' a_1 + \cdots + \mu_p' a_p$, $\mu_i' > 0$, means (Theorem 3.6) that $\{a_1, \cdots, a_p\}$ positively spans S. S is not E^n, by Theorem 3.2. Also (Theorem 3.4) $\{a_1, \cdots, a_p\}$ is positively independent.

The basis of (iii) has been chosen so as to include a_1. It remains to show it may be replaced by a minimal basis. But this is simply a matter of applying what is already proved to S. If $\{a_1, \cdots, a_p\}$ is not minimal, that is if the dimension of S is less than $p-1$, then some proper subset of $\{a_1, \cdots, a_p\}$ is a positive basis of a proper subspace of S; and so on. This process must end, since it decreases the dimension each time; when it ends, the minimal basis has been reached.

A subspace S such that those a_i in S are a basis for S will be called a *spanned subspace*; if the basis is minimal, a *minimal subspace.*

THEOREM 4.2. *Any spanned subspace is a linear sum of minimal subspaces, and conversely.*

Proof omitted.

In some ways the spanned subspaces are analogous to the coordinate subspaces determined by an arbitrary *linear basis.*

THEOREM 4.3. *For the spanned subspace S spanned by $\{a_1, \cdots, a_p\}$ the following are equivalent:*

(i) *S is minimal;*

(ii) *the positive relation satisfied by $a_1, \cdots, a_p$ is unique* (up to multiplication by a positive constant).

Proof. For any such S there is a strictly positive relation between $a_1, \cdots, a_p$, by Theorem 3.6. If S is not minimal it has a proper spanned subspace S', spanned, say, by $\{a_1, \cdots, a_q\}$, $q < p$; the strictly positive relation between $a_1, \cdots, a_q$ is a positive relation between $a_1, \cdots, a_r$ which is not strictly positive. Hence (ii) implies (i).

On the other hand, if (ii) fails then the dimension of S is less than $p-1$, so, by Theorem 4.1, (i) fails.

THEOREM 4.4. *Every linear relation between $a_1, \cdots, a_r$ is implied by*

the collection of strictly positive relations associated with the minimal subspaces according to Theorem 3.6 (iii).

It is convenient to delay the proof to § 5.

5. The convex cone associated with a basis. Instead of simply listing the members of the basis $\{a_1, \cdots, a_r\}$, one can describe it in another way. Consider Euclidean space E^r with a distinguished coordinate system, the unit coordinate vectors being $u_1, \cdots, u_r$, and let A be the linear transformation determined by the conditions $Au_i = a_i$, $i = 1, \cdots, r$ (cf. [5]). The elements of the null-space $N(A)$ of A correspond one-one to the linear relations holding between the a_i; and, since such relations determine the a_i to within affine isomorphisms in E^n, the geometrical properties of the basis can be described in terms of those of $N(A)$.

In particular, the positive relations between the a_i correspond to the elements of $N(A) \cap P^r$. Here P^r is the positive orthant of E^r, that is, the set of all elements of E^r without negative components.

The *facets* of the convex cone $N(A) \cap P^r$ are intersections of $N(A)$ with the facets of the convex cone P^r. (A q-1-dimensional *facet* of a q-dimensional cone is a maximal q-1-dimensional cone contained in its boundary; facets of lower dimension are defined by induction [6].) This follows from a general theorem, whose proof will only be sketched since it falls logically within the province of [6].

THEOREM 5.1. *Let C be a convex cone, S a linear subspace. Then the convex cones $S \cap F$, with F a facet of C, are all the facets of $S \cap C$.*

Proof Sketch. Every $S \cap F$ is proved to be a facet by Theorem 30 of [6]. Next any $x \in S \cap C$ is shown to lie in the relative interior of some $S \cap F$, and application of Theorem 35 of [6] completes the proof.

Now each facet Q of P^r is determined by the set of the u_i which it contains; in fact it is the set of all positive combinations of those u_i. $N(A) \cap Q$ is a facet of $N(A) \cap P^r$ if and only if some element of the relative interior of Q is in $N(A)$; that is, if and only if some strictly positive combination of the u_i in Q is mapped to 0 by A; that is, if and only if there is some strictly positive relation between the Au_i for which $u_i \in Q$; that is, if and only if those $Au_i = a_i$ for which $u_i \in Q$ span a spanned subspace.

THEOREM 5.2. *The lattice of spanned subspaces* (for the basis given by $A: E^r \to E^n$) *is isomorphic to the lattice of facets of $N(A) \cap P^r$ under the natural correspondence just exhibited.*

This has essentially been proved. That the spanned subspaces form a lattice when ordered by inclusion follows from Theorem 4.2. The facets of a cone are of course to be given their natural order in terms of incidence. Remaining details are left to the reader. Note that the lattice of spanned subspaces, though it is join-closed as a sub-partly-ordered set [1] of the lattice of all linear subspaces of E^n, has a different meet operation; that is, that the intersection of two spanned subspaces even when non-zero is not necessarily the largest spanned subspace contained in both. For a counterexample we clearly must find a case where it is not a spanned subspace at all. But in Example V the spanned subspace spanned by $a_1, \cdots, a_4$ and that spanned by $a_5, \cdots, a_8$ have a 1-dimensional intersection which does not even contain any a_i. (Cf. Theorem 6.6 (iii) below.) The source of such cases is seen easily from the lattice-isomorphism: facets of P^r such that $N(A)$ intersects the relative interior of each, but not that of their intersection.

Notice that *extreme rays of* $N(A) \cap P^r$ *correspond one-one to minimal subspaces* under the correspondence of the last theorem. They are the "points" (minimal non-null elements) of the respective lattices.

Note too that if Q corresponds to S then the elements of Q give all the positive relations involving only those a_i lying in S.

Proof of Theorem 4.4. Let $x_1, \cdots, x_s$ be non-zero vectors lying one in each extreme ray of $N(A) \cap P^r$. The conclusion of the theorem is equivalent to the statement that the linear combinations of $x_1, \cdots, x_s$ are all of $N(A)$. But their positive combinations are $N(A) \cap P^r$. Therefore it will suffice to show that an arbitrary $y \in N(A)$ is a difference of elements of $N(A) \cap P^r$. Only the case $y \notin P^r$ needs discussion.

By Theorem 3.6, some element z_1 of $N(A)$ lies in the relative interior of P^r. On the line joining y to z_1, namely $\{\lambda y + (1-\lambda)z_1;\ 0 \leqq \lambda \leqq 1\}$, there is some value $\lambda_0 > 0$ giving a point $z_2 \in P^r$, otherwise z_1 would be on the boundary: $z_2 = \lambda_0 y + (1-\lambda_0)z_1$. The equation can be solved for y, proving the theorem.

Theorem 5.2 showed that the "essential features" of the basis in the original space (more precisely, the incidence relations of the spanned subspaces and the membership of a_i in them) are given in a simple way by looking at $N(A) \cap P^r$. (Note that the intrinsic geometry of this cone does not suffice, but the way in which it is obtained as an intersection of a linear subspace with P^r is needed; but see § 6.) The proof just finished shows that all affine properties of the basis are so given. Further examination of $N(A) \cap P^r$ is suggested. Write hereafter $N_+(A) = N(A) \cap P^r$.

Let $x_1, \cdots, x_s$ be as in the proof of Theorem 4.4, and let X be the matrix whose elements are given by $X_{ij} = x_i \cdot u_j$. That is, X maps the unit vectors of Euclidean space of s dimensions onto the x_i, which give a frame of $N_+(A)$. Let a real matrix whose columns, considered as vectors, span the intersection of the positive orthant with a linear subspace, be called a *+-matrix*. It is proved in [3] that the transpose of a +-matrix is again a +-matrix, and the cones spanned by the columns of these +-matrices are intrinsically dual to each other. Let a +-matrix whose columns are positively independent, and whose rows are positively independent, be called a *+-canonical matrix*. In the present case, X is a +-matrix whose columns are positively independent. Hence it may be written (perhaps after changing the order of the rows) as $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, where X_1 is +-canonical and the rows of X_2 are positive combinations of those of X_1. Also, the cone associated with X_1 is intrinsically affine-isomorphic to that associated with X (namely $N_+(A)$), since each is dual to that associated with the transpose of X_1.

The principal task of this section is to give an independent characterization of those X which can occur.[2] The question is slightly reformulated, and the answer stated, in

THEOREM 5.3. *All positive bases of Euclidean spaces are obtained in the following way: Take Y any +-canonical matrix, Z any matrix whose rows are non-zero positive combinations of those of Y. Let*

$$X = \begin{pmatrix} Y \\ Y \\ Z \end{pmatrix}.$$

Use each column of X as the set of coefficients in a homogeneous linear relation $X_{i1}b_1 + \cdots + X_{ir}b_r = 0$. If $b_1, \cdots, b_r$ solve these relations and no further independent homogeneous linear relations, then $\{b_1, \cdots, b_r\}$ is a positive basis of the space it linearly spans.

Proof. There are two parts.

The proof that no bases are omitted has almost been completed already. With X_1 and X_2 as above, it must be shown, first that no row of X_2 is zero,

[2] They will be characterized only up to the following inessential changes: Any number of interchanges of columns, or multiplications of columns by a positive constant (this changes merely the ordering or the lengths of the x_i in P^r, therefore leaves $a_1, \cdots, a_r$ unaltered). Any number of interchanges of rows, or multiplications of a row by a positive constant (this changes the picture in E^r, but since it corresponds merely to changing the order or the lengths of the a_i it is not interesting either).

second that every row of X_1 is repeated among those of X_2. If a row of X were zero, some u_i would be orthogonal to all x_j, therefore to $N(A)$; but this contradicts the existence, proved above, of an element of $N(A)$ in the interior of P^r. If on the other hand, some row of X_1, say the first, is not repeated, then it is not a positive combination of the remaining rows ([6], Theorem 9). Using the first statement of Theorem 3.3, one shows that there is some set of s numbers $\lambda_1, \cdots, \lambda_s$ such that, if each equation $X_{i1}a_1 + \cdots + X_{ir}a_r = 0$ is multiplied by the appropriate λ_i and the results added, an equation is obtained in which only the coefficient of a_1 is positive. This means that a_1 is a positive combination of the other a_i, contradicting the positive independence of $\{a_1, \cdots, a_r\}$. This finishes half of the theorem. The proof of the other half follows.

Since every row of X is a positive combination of the remaining ones, there is no set of numbers $\lambda_1, \cdots, \lambda_s$ as above. (Use Theorem 3.3 again.) Therefore the relations imposed on $b_1, \cdots, b_r$ do not imply any positive dependence; therefore there is none, by hypothesis. It remains to show that $\{b_1, \cdots, b_r\}$ positively spans the space which it linearly spans. By Theorem 3.6, it is enough to produce a strictly positive relation between the b_i. This can be done by adding up the positive relations imposed on them. This completes the proof.

The notation of this theorem will be followed in the succeeding section.

6. Further details. The matrix Y alone determines some of the features of the basis: Since it determines the intrinsic geometry of $N_+(A)$, it determines the incidence relations of the spanned subspaces, by Theorem 5.2. Once Y has been fixed, adjoining of a new row to Z has the effect of increasing by one the dimension of E^n and the number of basis elements. This observation can be sharpened to the following.

THEOREM 6.1. *Let B be the projection of E^n on the orthogonal complement of a_r. If $\{Ba_1, \cdots, Ba_{r-1}\}$ is a basis of BE^n, then the matrix X associated with it by Theorem 5.3 is obtained from that associated with $\{a_1, \cdots, a_r\}$ by omitting the r-th row.*

B will be said to have *eliminated* a_r in this case.

Proof. $X_{i1}a_1 + \cdots + X_{ir}a_r = 0$, $i = 1, \cdots, s$, implies $X_{i1}Ba_1 + \cdots + X_{ir}Ba_r = 0$, which because of $Ba_r = 0$ gives $X_{i1}Ba_1 + \cdots + X_{i,r-1}Ba_{r-1} = 0$. Essentially all that needs to be added is that there can be no further independent relations satisfied by the Ba_j, $j = 1, \cdots, r$, or they could not linearly span an $n - 1$-dimensional space.

Notice the corollary that such a B carries spanned (minimal) subspaces into spanned (minimal) subspaces.

THEOREM 6.2. *a_r can be eliminated if and only if*[3] *the r-th row of X is in Z. Thus the necessary and sufficient condition for no a_i to be eliminable is that Z be absent.*

This follows from the last two theorems; the proof is omitted.

THEOREM 6.3. *In the case of a minimal basis, every a_i is eliminable, provided $n > 1$.*

Proof. In this case X has only one column, and (for $n > 1$) more than two rows.

This describes completely the relation of Z to the geometry of the basis. It is natural to ask what properties of the basis correspond to special properties of Y. First, consider the case where Y, and therefore also X, are reduced matrices:

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}.$$

THEOREM 6.4. *In case Z is absent, Y is a reduced matrix if and only if E^n is the direct sum of two non-zero spanned subspaces.*

Proof. Assume Z absent, Y reduced. Then the minimal subspaces fall into two classes, with no a_i in common: one corresponding to the columns of X_1, the other to those of X_2. Take the linear sum of each class. There result two spanned subspaces S_1 and S_2 whose linear sum is obviously E^n. It must be shown that $S_1 \cap S_2 = 0$. Suppose $b \in S_1 \cap S_2$, $b \neq 0$. Then b is a non-trivial positive combination of the a_i lying in S_i; $-b$ is a non-trivial positive combination of those lying in S_2; adding these two relations gives a positive dependence among the a_i, that is, an element $y \in N_+(A)$. It is easy to see that y is not a positive combination of the columns of $\begin{pmatrix} X_1 \\ 0 \end{pmatrix}$ and those of $\begin{pmatrix} 0 \\ X_2 \end{pmatrix}$; this is a contradiction, so $S_1 \cap S_2 = 0$.

For the other half of the theorem, let S_1 and S_2 be two such spanned subspaces. Every a_i is in one or the other; X will be shown to be reduced once it is shown that every minimal subspace is contained in either S_1 or S_2. Suppose $a_1, \cdots, a_p, a_q, \cdots, a_r$ span a minimal subspace S', where $a_1, \cdots, a_p$ are in S_1 and $a_q, \cdots, a_r$ are in S_2. For some $\mu_i > 0$,

[3] With the assumed proviso that if a row of X is identical or proportional to one in Z the corresponding a_i can be eliminated; this is again a mere matter of the order in which the rows are written, cf. fn. 2. For a trivial example of this, see Example III.

$$\mu_1 a_1 + \cdots + \mu_p a_p + \mu_q a_q + \cdots + \mu_r a_r = 0.$$

Now $\mu_1 a_1 + \cdots + \mu_p a_p \in S_1$ and $\mu_q a_q + \cdots + \mu_r a_r \in S_2$ are not zero, otherwise S' would not be minimal. But they add to zero, contradicting the linear independence of S_1 and S_2. The theorem is proved.

The extreme case here is where Y is a diagonal matrix. Notice that Y is in fact always diagonal for $n = 3$ (Examples I-IV). This is also true for $n = 4$, as will be proved in Theorem 6.8; but for $n \geqq 5$ there exist bases with non-trivial Y, such as Example V. Before diagonal Y are discussed, it is convenient to mention this simple fact:

THEOREM 6.5. *Let $a_1, \cdots, a_p$ be all the a_i lying in the minimal subspaces corresponding to $x_1, \cdots, x_q$. Let S, of dimension m, be the linear subspace of E^n linearly spanned by $a_1, \cdots, a_p$. Let the dimension of the smallest facet of $N_+(A)$ containing $x_1, \cdots, x_q$ be t. Then $m = p - t$.*

Proof. S is a spanned subspace, that one in fact which corresponds to Q by Theorem 5.2. There are no linear relations among $a_1, \cdots, a_p$ which are linearly independent of those given by elements of Q. Now if A is considered as a function on $A^{-1}S$ to S, its domain has dimension p, its null-space has dimension t, and its range has dimension m, which proves the theorem.

A particularly evident special case is $m = n$, $p = r$. Then t is the dimension of $N(A)$, that is, rank (X), which in turn is equal to rank (Y); so $n = r - \text{rank } (Y)$.

THEOREM 6.6. *The following are equivalent:*

(i) *Y is a diagonal matrix;*

(ii) *the number of minimal subspaces equals the number of basis elements minus the dimensionality of the space: $s = r - n$.*

A necessary condition is

(iii) *the intersection of any two minimal subspaces is linearly spanned by those a_i it contains.*

Proof. Since Y has s columns, $s \geqq \text{rank } (Y) = r - n$. The equivalence of (i) and (ii) therefore reduces to the following lemma:

LEMMA. *A $+$-canonical matrix is diagonal if and only if the rank equals the number of columns.*

Proof. Let the matrix again be called Y, and have s columns, all linearly independent. The transpose Y' is also $+$-canonical. Its columns are points

in E^s giving a frame of the intersection of a linear subspace with the positive orthant P^s. The number of linearly independent columns of Y' is s, hence the linear subspace is all E^s, hence Y is (perhaps after a rearrangement of rows) diagonal. The other half of the lemma is trivial.

It remains to consider condition (iii) of the theorem. The a_i common to two minimal subspaces are linearly independent in general, by Theorem 4. 1. The only possibility of failure of (iii) is that the dimension of the intersection may be greater than the number of common a_i.

Let the two minimal subspaces, corresponding to x_1 and x_2, be of dimensions m_1 and m_2; let the common a_i be p' in number; let the linear span of the subspaces have dimension m and contain p of the a_i. Clearly $p = (m_1 + 1) + (m_2 + 1) - p'$. The dimension of the intersection is $m_1 + m_2 - m$. For (iii) to fail for x_1, x_2 is for $m_1 + m_2 - m > p'$ to hold; that is, for $p - m > 2$ to hold; that is, for $t > 2$ to hold, t being as in Theorem 6. 5; that is, for the two extreme vectors x_1, x_2 of $N_+(A)$ not to lie in a single 2-facet of $N_+(A)$.

Thus (iii) cannot fail if $N_+(A)$ is formed from a simplex by projection through a vertex, that is if Y is diagonal. Theorem 6. 6 is proved.

Note that (iii) is *not* sufficient for (i); a counterexample is given by Example VII.

It is clear that, for any n, E^n admits a basis consisting of the elements of a linear basis together with their negatives. Such a basis might be called *maximal*, in contrast to minimal bases. The following theorem (which in a sense complements Theorem 3. 8) justifies the terminology. (Theorem 6. 7 is essentially due to Robinson and Blumenthal; see [2], sections 76 and 120. It is seen to follow more naturally from the present strictly algebraic theory than from their metric approach.)

THEOREM 6. 7. *$r \leqq 2n$, and the following are equivalent:*

(i) $r = 2n$;

(ii) *Y is diagonal and Z is absent;*

(iii) *every minimal subspace is one-dimensional, so that the basis is maximal.*

Proof. As above, $n = r - \text{rank } (Y)$. Now rank $(Y) \leqq s'$, the number of rows of Y, with equality only if Y is diagonal, by the dual of the Lemma above. Also $2s' \leqq r$, with equality only if Z is absent, by the basic Theorem 5. 3. Therefore $2n \geqq 2r - 2s' \geqq r$, with equality only if Y is diagonal and Z is absent. It remains to show that (ii) implies (iii), (iii) implies (i), but these are trivial.

Here is one more application to a special case:

THEOREM 6.8. *For $n \leqq 4$, every basis gives diagonal Y.*

Proof. The number s' of rows of Y satisfies $s' \leqq r/2 \leqq n \leqq 4$. Any non-diagonal $+$-canonical matrix must have at least four rows (because three vectors in a single orthant are either linearly independent or positively dependent, evidently); so we need consider only $s' = 4$, $r = 8$, $n = 4$. But in that case also Y must be diagonal, for rank $(Y) = r - n = 4 = s'$.

7. Examples. In each of the examples I-VII below, A and X are as in § 5: the columns of A are $a_1, \cdots, a_r$, the rows of X' (the transpose of X) are $x_1, \cdots, x_s$.

I. $$A = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \qquad X' = (1 \quad 1 \quad 1 \quad 1).$$

II. $$A = \begin{bmatrix} 1 & 1 & 1 & 1 & -2 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix}, \qquad X' = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

III. $$A = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \qquad X' = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

IV. $$A = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix},$$

$$X' = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

V. $$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix},$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

In the last two examples, only Y' is given, Y being as in Theorem 5.3.

VI. $$Y' = \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 3 & 4 & 0 & 0 & 1 \\ 1 & 0 & 0 & 4 & 3 \end{bmatrix}.$$

VII. $$Y' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

ANN ARBOR, MICHIGAN.

REFERENCES.

[1] G. Birkhoff, *Lattice theory*, Rev. ed., American Mathematical Society Colloquium Publications, No. XXV, New York, 1948.

[2] L. M. Blumenthal, *Theory and applications of distance geometry*, Oxford, 1953.

[3] C. Davis, "The intersection of a linear subspace with the positive orthant," *Michigan Mathematical Journal*, vol. 1 (1952), pp. 163-168; also, vol. 2 (1953), pp. 23-25.

[4] J. Favard, "Sur les corps convexes," *Journal de mathématiques pures et appliqués*, vol. 12 (1933), pp. 219-282.

[5] D. Gale, "Convex polyhedral cones and linear inequalities," Chap. XVII of Cowles Commission Monograph No. 13, *Activity analysis of production and allocation*, ed. T. C. Koopmans, Wiley, New York, 1951.

[6] M. Gerstenhaber, "Theory of convex polyhedral cones," Chap. XVIII of Cowles Commission Monograph No. 13, *Activity analysis of production and allocation*, ed. T. C. Koopmans, Wiley, New York, 1951.

[7] T. Motzkin, *Beiträge zur Theorie der linearen Ungleichungen*, Basel Doctoral Thesis, 1933. Transl. D. R. Fulkerson as RAND report T-22, 1952.

[8] E. Steinitz, "Polyeder und Raumeinteilungen," Encyklopädie der Mathematischen Wissenschaften III AB 12.

MODAL OPERATORS, EQUIVALENCE RELATIONS, AND PROJECTIVE ALGEBRAS.* [1]

By CHANDLER DAVIS.

---

1. Introduction. In modal logic, as initiated by C. I. Lewis in 1918, one applies to propositions $a, b, \cdots$ not only the ordinary Boolean operations $\cap$, $\cup$, and $'$, but also a modal operator which for present purposes will be written C. This is a constant symbol with the following interpretation: If a is thought of as assertion the truth of a proposition, then Ca is thought of as asserting its *possibility*; similarly, $(Ca')'$ asserts its *necessary truth.*

Depending on the sense put on the words "possible" and "necessary," different properties will be assumed for C applied to propositional variables [12]. These properties may be expressed as identities, such as those below. Waiving the logical questions, some of these sets of identities may be studied as definitions of kinds of operators on Boolean algebras. Particularly interesting are $S4$ operators, satisfying

$$Ca \supseteq a, \tag{1.1}$$

$$CO = O, \tag{1.2}$$

$$CCa = Ca, \tag{1.3}$$

$$C(a \cup b) = Ca \cup Cb. \tag{1.4}$$

Here (1.1)-(1.3) are, with the condition of "isotoneness"

$$C(a \cup b) \supseteq Ca, \tag{1.4'}$$

just the assumptions made on "closure operators" on an arbitrary lattice [1, 19]; while all of (1.1)-(1.4) are satisfied if C is the operation of topological closure in a topological space, a and b being subsets and the Boolean operations being set operations. Boolean algebras with an $S4$ operator have been studied as generalizations of topological spaces [16].

* Received September 12, 1953.

[1] This material is essentially excerpted from the author's thesis [5]. The work was done without knowledge of the earlier work of Chin and Tarski (see [4]), which it presumably overlaps a good deal. It is unfortunate that their investigation has never been published.

$S5$ operators satisfy, beside (1.1)-(1.4),

(1.5) $\qquad a \cap Cb = O \quad \text{implies} \quad Ca \cap Cb = O.$

In the case of a topological space, the condition (1.5) on the closure operator implies that every open set is closed; in fact, (1.5) with (1.1) implies in general

(1.5′) $\qquad C(Ca)' = (Ca)'.$

Boolean algebras with an $S5$ operator do not therefore generalize any topologically interesting spaces, but they do provide generalizations of the simpler and "more set-theoretical" notion of equivalence relation.

This will be discussed more fully, together with other preliminary remarks on $S5$ operators, in Section 2. In Section 3 is a discussion of the reason for considering Boolean algebras with several $S5$ operators. The remainder of the paper gives theorems on the structure of these systems. They are shown to be in a rough sense "coordinatizable" by deriving them from the projective algebras of Everett and Ulam (Sections 4-5); only the two-operator case is treated here. The relation of these systems to the first-order functional calculus of logic is stated in Section 6.

The terminology to be used for Boolean algebras is as follows. "Meet" means "$\cap$," "join" means "$\cup$." The relation $a = a \cup b$, i.e., $a \supseteq b$, will be called "a is above b" or "b is below a." I and O are the "unit" and "null" elements respectively. A prime denotes complement if it follows an element of a Boolean algebra, though it may have other meanings in other connections. If $\mathcal{A}$ is a *set of elements* of a Boolean algebra, $\cup \mathcal{A}$ will be the join of all elements in the set.

2. Properties of $S5$ operators. First of all, the following more economical set of conditions will be convenient:

LEMMA 2.1. *If $\mathfrak{B}$ is a Boolean algebra, any function C on $\mathfrak{B}$ into $\mathfrak{B}$ is an $S5$ operator provided it satisfies*

(1.1) $\qquad Ca \supseteq a,$

(1.4′) $\qquad C(a \cup b) \supseteq Ca,$

(1.5′) $\qquad C(Ca)' = (Ca)'.$

Proof. (1.4′) is a consequence of (1.4) alone. As already mentioned, (1.1) and (1.5) give (1.5′): $Ca \cap (Ca)' = O$, hence by (1.5), $Ca \cap C(Ca)' = O$, that is, $C(Ca)' \subseteq (Ca)'$, and the inequality cannot be strict because of (1.1).

The other half is to show that (1.1), (1.4′), (1.5′) imply (1.1)-(1.5). Now (1.1) implies $CI = I$, so, by (1.5′), $CO = C(CI)' = (CI)' = O$; (1.2) is proved. Next, (1.3) results from (1.5′) alone: $CCa = C(Ca)'' = C(C(Ca)')' = (C(Ca)')' = (Ca)'' = Ca$. The proof of (1.5) is also easy: Suppose $a \cap Cb = O$. This is equivalent to $a \subseteq (Cb)'$, so that by (1.4′) and (1.5′), $Ca \subseteq C(Cb)' = (Cb)'$, that is, $Ca \cap Cb = O$. (1.5) is proved. Only (1.4) remains.

In the usual terminology, call any fixpoint of C "closed"; the closed elements are those of the form Ca, by (1.3). Well-known arguments show that

(2.1) $\quad a \subseteq Cb$ implies $Ca \subseteq Cb$,

(2.2) $\quad a$ and b closed implies $a \cap b$ closed.

Indeed, $a \subseteq Cb$ gives $Ca \subseteq CCb = Cb$ by (1.4′) and (1.3). And for a and b closed, (1.4′) gives $C(a \cap b) \subseteq Ca \cap Cb = a \cap b$; the inequality cannot be strict because of (1.1), so (2.2) is also proved.

Complements and meets of closed elements are closed, by (1.5′) and (2.2) respectively, so $C\mathfrak{B}$ is a subalgebra and

(2.3) $\quad a$ and b closed implies $a \cup b$ closed.

The proof of (1.4) is now easy. By (1.1), $a \cup b \subseteq Ca \cup Cb$; by (2.3), $Ca \cup Cb$ is closed; by (2.1), $C(a \cup b) \subseteq Ca \cup Cb$. But the reverse inequality is obvious from (1.4′). The proof is complete.

It is easily shown that (1.1), (1.4′), (1.5′) are independent conditions if $\mathfrak{B}$ has at least eight elements.

The following theorems make precise the sense in which $S5$ operators generalize equivalence relations.

First the relation to equivalence relations is stated in the case where $\mathfrak{B}$ is a full Boolean algebra, that is, isomorphic to the Boolean algebra of all subsets of some set:

THEOREM 2.1. *Let $\mathfrak{B}$ be a full Boolean algebra. For each equivalence relation ϕ on the points of $\mathfrak{B}$, let C_ϕ be the operator defined as follows: for $a \in \mathfrak{B}$, $C_\phi a$ is the join of all points of $\mathfrak{B}$ which are ϕ-equivalent to any point below a. Then C_ϕ is $S5$; and the correspondence $\phi \to C_\phi$ is a lattice isomorphism of the partition lattice $P(\mathfrak{B})$ of all ϕ onto the lattice of all $S5$ operators on $\mathfrak{B}$. Also $C\mathfrak{B}$ is always a subalgebra of $\mathfrak{B}$ complete in $\mathfrak{B}$ and a full Boolean algebra; and the correspondence $\phi \to C_\phi\mathfrak{B}$ is a lattice dual isomorphism of $P(\mathfrak{B})$ onto the lattice of all such subalgebras.*

Explanations. $S5$ operators are ordered as functions into a poset [1], namely, $C_1 \supseteq C_2$ means that, for all $a \in \mathfrak{B}$, $C_1 a \supseteq C_2 a$. Equivalence relations on a set of points $p, q, \cdots$ are ordered by defining $\phi \supseteq \psi$ to mean that $p \psi q$ implies $p \phi q$. Subalgebras of $\mathfrak{B}$ are ordered as usual: by inclusion. A point of a Boolean algebra is a minimal element strictly above O. The join referred to in the second sentence of the theorem is of course the join operation of $\mathfrak{B}$, not set union.

This theorem is not really new (see [16, 17]; also [18, 19]); also the ideas involved are perhaps made sufficiently clear by the remarks in Section 1. Accordingly, the proof is omitted; it is given in [5].

Next, a representation theorem relates the general case to the special case above.

THEOREM 2.2. *If C is an $S5$ operator on the Boolean algebra $\mathfrak{B}$, and if $\mathfrak{B}'$ is the full Boolean algebra in which $\mathfrak{B}$ is imbedded by the Stone representation theorem, then there is an $S5$ operator C' on $\mathfrak{B}'$ which is an extension of C. If several $S5$ operators are extended to $\mathfrak{B}'$ in this way, the correspondence is order-preserving in both directions.*

This is a special case of a representation theorem of Jónsson and Tarski [10], but the proof is particularly simple in this case.[2]

For p a point of $\mathfrak{B}'$, define $C'p = \cap \{a; a \supseteq p, a \in C\mathfrak{B}\}$; define further $CO = O$. For other $b \in \mathfrak{B}'$, define $C'b$ as $\cup \{C'p; p \subseteq b, p \text{ a point}\}$. Then C' may be proved to be an $S5$ operator by Lemma 2.1, or as follows.

Define $p \phi q$, for points p and q of $\mathfrak{B}'$, to mean that $p \subseteq Ca$ implies $q \subseteq Ca$ for $a \in \mathfrak{B}$. The relation ϕ is obviously reflexive and transitive. To verify that it is symmetric, take $p \phi q$ and $p \not\subseteq Ca$; then, by (1.5'), $p \subseteq (Ca)' = C(Ca)'$, so $q \subseteq (Ca)'$, $q \not\subseteq Ca$. Hence ϕ is an equivalence relation. The $S5$ operator C_ϕ corresponding to it according to Theorem 2.1 is identical with C', by definitions.

C' must be shown to agree with C on $\mathfrak{B}$. The definition shows that if a is closed under C it is closed under C'. From this it follows that, for all $b \in \mathfrak{B}$, $C'b \subseteq Cb$. The reverse inequality is more difficult.

Take any $b \in \mathfrak{B}$; $C'b = \cup \{C'q; q \subseteq b, q \text{ a point}\}$; and each such $C'q$ is the join of an equivalence class under ϕ. Hence to show that $Cb \subseteq C'b$ it is enough to show, for any point $p \subseteq Cb$, the existence of a point $q \subseteq b$ with $p \phi q$.

Consider the set $\alpha = \{b \cap d; d \supseteq p, d \in C\mathfrak{B}\}$. First, α contains with

[2] The referee contributed to its simplicity.

any two elements their meet: if $b \cap d \in \alpha$, $b \cap \bar{d} \in \alpha$, then, because $d \cap \bar{d} \supseteq p$ and $d \cap \bar{d} \in C\mathfrak{B}$, $b \cap d \cap b \cap \bar{d} \in \alpha$. Next, α does not contain O: for $b \cap d = O$ and $d \in C\mathfrak{B}$ would imply $b \subseteq d' \in C\mathfrak{B}$, $d' \supseteq Cb \supseteq p$, which would forbid $d \supseteq p$. A standard argument shows there is some point q of $\mathfrak{B}'$ which is below every element of α; in particular, $q \subseteq b$. Also $p\phi q$, by the definition of α. Therefore $Cb \subseteq C'b$.

This proves the first statement of the theorem.

Now let C_1' and C_2' be the $S5$ operators on $\mathfrak{B}'$ obtained in this way from the $S5$ operators C_1 and C_2 on $\mathfrak{B}$. That $C_1' \subseteq C_2'$ implies $C_1 \subseteq C_2$ is obvious. The converse follows almost as easily: If $C_1 \subseteq C_2$, then $C_2\mathfrak{B}$ is contained in $C_1\mathfrak{B}$, so by definition $C_1'p \subseteq C_2'p$ for p a point of $\mathfrak{B}'$, so $C_1'b \subseteq C_2'b$ for all $b \in \mathfrak{B}'$, which was to be shown.

3. Discussion of $S5$ operators. All the considerations of this paper were motivated by an interpretation of the notion of possibility, therefore of modal operators, which will now be set forth.[3] It is related to the ideas of Wajsberg [22], Carnap [2, 3], and McKinsey [14].

Let us consider a logical system as a way of representing the relationships between statements which may correctly or incorrectly be made about the universe. (The "universe" here may be any "isolated" physical system or conceptual system.) No two such statements need be considered as distinct unless it is believed that one of them *might* hold without the other's doing so. In other words, let us consider a statement to assert something only insofar as it distinguishes a state of affairs which it asserts from an otherwise conceivable state of affairs which it denies. This point of view leads naturally to limiting oneself to propositions "in extension": To begin with, one imagines a set of "possible" alternate universes. Thereafter, a proposition a is considered "in extension," that is, it is identified with the set of alternate universes in which it holds. (Each proposition considered is to be one which asserts something about the universe: one, therefore, which can assert something about each alternate universe, and can hold or not hold in each alternate universe. It is not to be a proposition about sets of alternate universes; nor yet a proposition about propositions.) Now if the proposition asserted by a holds in some alternate universe one can say that the proposition asserted by Ca, "it is possible that a," holds in all the alternate universes.

Now in the formal system constructed following these ideas, logical

[3] The discussion of logic in this section, as in § 1, is informal. Nowhere in this paper is there a formal discussion of logical systems generalizing Lewis's $S5$, though the mathematical systems treated would be matrices for such systems.

connectives between propositions will correspond to operations on the elements of a complemented lattice of subsets of the set of all alternate universes. This lattice might be taken to be the Boolean algebra of *all* subsets, as is done by Carnap [3], who distingiushes each of his alternate universes initially by the subset of a given set of atomic propositions which hold on it. Here, the lattice will be taken to be some field of sets, therefore a Boolean algebra (although it is not at all clear that other lattices are uninteresting in this connection); it will not be assumed full.

The operator C on this Boolean algebra which was discussed above gives $CO = O$, and for $a \neq O$, $Ca = I$. It is an $S5$ operator—that which corresponds, if the Boolean algebra is full, to an equivalence relation with only one equivalence class.

Non-trivial $S5$ operators, or more than one $S5$ operator, might appear in the same context. An early argument for the introduction of two modal operators [6, 21] need not be discussed here (see [11]). A strong argument, however, is made by Weyl [23] for considering more than one "degree of possibility." He gives the following example. Suppose a train leaves Seattle at 9:00 a. m. Pacific Time. It is physically impossible for it to arrive in Chicago one microsecond later. It is technically impossible, though physically possible, for it to arrive in Chicago at 10:00 a. m. Pacific Time the same day. Weyl's example illustrates the desirability of a totally ordered set of "degrees of possibility"[4]; it can be amplified as follows to illustrate the case for a partly ordered set. It may be technically possible for the train to arrive at 9:00 a. m. two days later, but impossible because no track is available for it to enter the Chicago station at that time; the technically impossible arrival time of 10:00 on the day it started might however be possible by the latter criterion. Thus we have two "degrees of possibility" neither of which implies the other. The example is artificial, and a different one will serve better.

Let the alternate universes be "all" division rings. Assume for simplicity that the Boolean algebra being considered is that of *all* sets of alternate universes; that is, that there is, among those propositions being considered, a proposition distinguishing each of the division rings. Let $C_1 a$ assert that the proposition asserted by a is possible as far as commutativity or non-commutativity is concerned. (Equivalently one could say, let C_1 be C_ϕ of Theorem 2. 1, taking for ϕ the equivalence relation having two equivalence classes: the commutative division rings, the non-commutative ones.) Thus if a asserts,

[4] Possibility operators to be ordered as functions into the Boolean algebra of propositions, see § 2.

"This division ring has two elements," then C_1a is true for all fields (all members of the first equivalence class) and for no other division rings. Let C_2a assert that the proposition asserted by a is possible as far as finiteness or non-finiteness of the division ring is concerned. If a asserts, "The identity of this division ring has infinite order," then C_2a is true for any division ring with an infinite number of elements and for no others; and so forth. C_1 and C_2 are again $S5$ operators, and neither $C_1 \subseteq C_2$ nor $C_1 \supseteq C_2$.

Return for a moment to the first, relatively trivial interpretation of possibility above. It would have been equally reasonable to replace the propositions a by propositional functions $F_a(\cdot)$ in the following way. For each proposition a about the universe, let $F_a(x)$ assert that a holds in the alternate universe x. (For fixed x, $F_a(x)$ is not an assertion of the same form as a, so that expressions involving the $F_a(\cdot)$ cannot meaningfully be combined with expressions involving the a by logical connectives.) The association of a to $F_a(\cdot)$ is of course a Boolean isomorphism. To what does the $S5$ operator C described above correspond? By its definition, $Ca = I$ if $(\exists y)F_a(y)$ is true, $Ca = O$ otherwise. If one is willing to regard the proposition $(\exists y)F_a(y)$ as a propositional function (whose variable is not written and is without effect on the truth-value), then the conclusion may be simply stated: The operator C on propositions corresponds to existential quantification on the associated propositional functions. This correspondence between modal propositional calculus and the ordinary "einstellige Prädikatenkalkul" was mentioned but not discussed by Wajsberg [22].

So much for the case of one trivial $S5$ operator. The introduction of more than one modal operator naturally suggests the extension from the calculus of propositional functions of one variable to the first-order functional calculus, in which propositional functions may have more than one argument. In fact, there is a close relationship between the two situations; this relationship is the subject of the following sections.

4. *S5* operators and projective algebras. The projective algebras of Everett and Ulam [7] were intended as a preliminary attempt at supplying a model for the first-order functional calculus similar to that supplied for propositional calculus by Boolean algebra. The definition is as follows.[5]

A projective algebra is a Boolean algebra $\mathfrak{Q}$ (whose unit element will be written "i"), with a distinguished point p_0; with unary operations ("projections") $a \to a_x$ and $a \to a_y$; and with a binary operation $a, b \to a \square b$ defined for $a \subseteq i_x$ and $b \subseteq i_y$; subject to the following laws:

[5] See also [4].

P1. $(a \cup b)_x = a_x \cup b_x$; $(a \cup b)_y = a_y \cup b_y$.

P2. $i_{xy} = i_{yx} = p_0$.

P3. $a = O$, $a_x = O$, and $a_y = O$ are all equivalent.

P4. $a_{xx} = a_x$, $a_{yy} = a_y$.

P5. Let $O \subset a \subseteq i_x$, $O \subset b \subseteq i_y$. Then $(a \square b)_x = a$, $(a \square b)_y = b$; also, for any $c \in \mathfrak{Q}$, $c_x = a$ and $c_y = b$ imply $c \subseteq (a \square b)$.

P6. $i_x \square p_0 = i_x$; $p_0 \square i_y = i_y$.

P7. $(a \cup \bar{a}) \square i_y = (a \square i_y) \cup (\bar{a} \square i_y)$; $i_x \square (b \cup \bar{b}) = (i_x \square b) \cup (i_x \square \bar{b})$.

D1. $a \square 0 = 0 \square b = 0$.

The typical case motivating these conditions is the projective algebra of subsets of a direct product, in which $\mathfrak{Q}$ is a Boolean algebra of some subsets of a Cartesian product of a set by itself, and where the projections $a \to a_x$ and $a \to a_y$ are the projections of a set in $\mathfrak{Q}$ onto the respective axes. Every projective algebra is isomorphic to a projective algebra of this sort [7, 15], but not necessarily to a projective algebra of *all* subsets of a direct product.

Every projective algebra $\mathfrak{Q}$ can be made into a Boolean algebra with two $S5$ operators in a natural way. Namely, $S5$ operators on $\mathfrak{Q}$ as a Boolean algebra are given by defining

(A) $$C_1 a = a_x \square i_y, \qquad C_2 a = i_x \square a_y.$$

This result is a special case (where e is taken equal to i) of Theorem 4.1 which follows.[6]

THEOREM 4.1. *Let $\mathfrak{Q}$ be a projective algebra and e a fixed element of $\mathfrak{Q}$. On the Boolean algebra of elements below e (complement being defined as relative complement) the following are $S5$ operators:*

(B) $$C_1 a = (a_x \square i_y) \cap e, \qquad C_2 a = (i_x \square a_y) \cap e.$$

THEOREM 4.2. *Every Boolean algebra with two $S5$ operators is obtained in this way.*

Proof of Theorem 4.1. It is enough to give the argument for C_1, since the proof for C_2 would go just the same. The conditions of Lemma 2.1 must be verified; accordingly, it will be shown that

(4.1) $$(a_x \square i_y) \cap e \supseteq a;$$

[6] Cf. remark at end of § 5.

(4.4) $((a \cup b)_x \square i_y) \cap e = ((a_x \square i_y) \cup (b_x \square i_y)) \cap e;$

(4.5′) $C_1((C_1a)' \cap e) = (C_1a)' \cap e.$

Here (1.5′) takes the form (4.5′) because the prime is still being used to denote the complement in $\mathfrak{Q}$ (not the relative complement); (4.4) is the statement of (1.4), which is stronger than (1.4′).

These identities can be verified using the representation theorem for projective algebras, or formally, as follows. In the proof, results from [7] will be cited by number, without quoting them.

First, if $a = O$ then (4.1) becomes trivial and (4.4) follows from P3, D1. If none of the elements involved is O, one has, by C1, C9, and C10, $(a_x \square i_y) \cap e \supseteq (a_x \square a_y) \cap e \supseteq a$, proving (4.1); while C19 gives (4.4).

If $a = O$, (4.5′) reduces to $C_1e = e$, which follows from (B) and (4.1). It remains to prove (4.5′) for $O \subset a \subseteq e$. Now

$$\begin{aligned}
(C_1a)' \cap e &= (a_x \square i_y)' \cap e && \text{(obvious from (B))}\\
&= ([(a_x' \cap i_x) \square i_y] \cup [i_x \square (i_y' \cap i_y)]) \cap e && \text{(by C22)}\\
&= ((a_x' \cap i_x) \square i_y) \cap e && \text{(by D1).}
\end{aligned}$$

Also

$$\begin{aligned}
C_1((C_1a)' \cap e) &= ([(a_x \square i_y)' \cap e]_x \square i_y) \cap e && \text{(by (B))}\\
&\subseteq ((([(a_x \square i_y)']_x \cap e_x) \square i_y) \cap e && \text{(by C4, C9)}\\
&= ([(a_x \square i_y)']_x \square i_y) \cap (e_x \square i_y) \cap e && \text{(by C11)}\\
&= ([(a_x \square i_y)']_x \square i_y) \cap e && \text{(by P5, C9)}\\
&= ([(a_x' \cap i_x) \square i_y]_x \square i_y) \cap e && \text{(see above)}\\
&= ((a_x' \cap i_x) \square i_y) \cap e && \text{(by P5).}
\end{aligned}$$

Hence "$\subseteq$" holds in (4.5′). But the reverse inequality is immediate from (4.1). The proof is finished.

5. Continuation. *Proof of Theorem* 4.2. Imbed the given Boolean algebra $\mathfrak{B}$ in $\mathfrak{B}'$ by Theorem 2.2; let the extensions of the given $S5$ operators C_1 and C_2 be written simply C_1 and C_2. The proof is in three parts: construction of a certain full projective algebra $\mathfrak{Q}'$; construction of a certain sub-projective-algebra $\mathfrak{Q}$ of $\mathfrak{Q}'$; and showing $\mathfrak{B}$ can be obtained from $\mathfrak{Q}$ by the method of Theorem 4.1.

Elements of $\mathfrak{B}'$ will be writen as before (in particular, "p" and "q" will denote points); lower-case Greek letters will denote elements of $\mathfrak{Q}'$ (thus its unit element is written "ι").

Part I. The points π of $\mathfrak{Q}'$ are to be as follows:

(5.1) for every point p of $\mathfrak{B}'$, a point $[p]$;

(5.2) for every point C_1p of $C_1\mathfrak{B}'$ and point C_2q of $C_2\mathfrak{B}'$, a point $[C_1p, C_2q]$;

(5.3) for every point C_1p of $C_1\mathfrak{B}'$, a point $[C_1p]$;

(5.4) for every point C_2q of $C_2\mathfrak{B}'$, a point $[C_2q]$;

(5.5) an additional point π_0.

Here it is clear that the same element r of $\mathfrak{B}'$ may give rise to more than one point of $\mathfrak{Q}'$—e. g., if r is a point both of $\mathfrak{B}'$ and of $C_1\mathfrak{B}'$. In such a case the notations $[r]_1$, for the point of the form (5.1), and $[r]_3$, for the point of the form (5.3), could be used to distinguish the two; however, the arguments which follow will be sufficiently clear without doing this. The point π_0 will be the distinguished point (§ 4) of $\mathfrak{Q}'$.

Let $\mathfrak{Q}'$ be the full Boolean algebra generated by the points (5.1)-(5.5). Denote by χ_n the join of all points of the form $(5.n)$, $n = 1, 2, 3, 4, 5$.

The projections of $\mathfrak{Q}'$ will be defined first for points π.

π_x:	$[p]_x = [C_1p]$	π_y:	$[p]_y = [C_2p]$
	$[C_1p, C_2q]_x = [C_1p]$		$[C_1p, C_2q]_y = [C_2q]$
	$[C_1p]_x = [C_1p]$		$[C_1p]_y = \pi_0$
	$[C_2p]_x = \pi_0$		$[C_2p]_y = [C_2p]$
	$(\pi_0)_x = \pi_0$;		$(\pi_0)_y = \pi_0$.

Next, for arbitrary $\alpha \in \mathfrak{Q}'$, define $\alpha_x = \cup \{\pi_x; \pi \subseteq \alpha, \pi \text{ a point}\}$; and α_y similarly.

Of the conditions (P1-P7 and D1) which must be checked to prove $\mathfrak{Q}'$ a projective algebra, P1, P3, and P4 are already clear from the definitions.

Now $\iota_x = \chi_3 \cup \chi_5$, $\iota_y = \chi_4 \cup \chi_5$, in the notations above. Recognizing this, one verifies P2 immediately.

Also for $\alpha \in \mathfrak{Q}'$, $\alpha \subseteq \iota_x$ if and only if α is a join of points of the forms (5.3) and (5.5), with a similar condition for $\beta \subseteq \iota_y$. For such α and β, the operation "$\square$" must be defined. First, $\alpha \square O = O \square \beta = O$ may be assumed; this is D1. Next, for $O \subset \alpha \subseteq \iota_x$ and $O \subset \beta \subseteq \iota_y$, let $\alpha \square \beta = \cup \{\pi; \pi_x \subseteq \alpha, \pi_y \subseteq \beta, \pi \text{ a point}\}$. This makes P6 and P7 obvious.

P5 must be checked. As a sample, here is the demonstration that $\alpha \subseteq (\alpha \square \beta)_x$. Consider any point $\pi \subseteq \alpha$. If $\pi = \pi_0$, then $\pi = (\pi \square \rho)_x$, where $\rho = \pi \square \rho$ is any point below β. (Remember $\beta \neq O$.) The other case

is $\pi = [C_1 p]$: then either $\pi = [C_1 p, C_2 q]_x$, with $[C_2 q] = [C_1 p, C_2 q]_y \subseteq \beta$; or, if $\beta = \pi_0$ one has $\pi = \pi_x$, $\pi_0 = \pi_y$. (It might also happen that $\pi = [p]_x$, with $[p]_y \subseteq \beta$, but this can not be guaranteed; this was of course the reason for introducing the points (5.2) originally.) In either case, $\pi \subseteq (\alpha \square \beta)_x$. Therefore $\alpha \subseteq (\alpha \square \beta)_x$.

It has been proved that $\mathfrak{Q}'$ is a projective algebra.

Part II. Next the subset $\mathfrak{Q}$ of $\mathfrak{Q}'$ will be defined. Here $\mathfrak{A}, \mathfrak{A}_1, \mathfrak{A}_2$ represent sets of *points* of (respectively) $\mathfrak{B}'$, $C_1\mathfrak{B}'$, $C_2\mathfrak{B}'$.

(5.6) If $\cup \mathfrak{A} \varepsilon \mathfrak{B}$, then $\cup\{[p]; p \varepsilon \mathfrak{A}\} \varepsilon \mathfrak{Q}$;

(5.7) if $\cup \mathfrak{A}_1 \varepsilon C_1\mathfrak{B}$, $\cup \mathfrak{A}_2 \varepsilon C_2\mathfrak{B}$,

then $\cup \{[C_1 p, C_2 q]; C_1 p \varepsilon \mathfrak{A}_1, C_2 q \varepsilon \mathfrak{A}_2\} \varepsilon \mathfrak{Q}$;

(5.8) if $\cup \mathfrak{A}_1 \varepsilon C_1\mathfrak{B}$, then $\cup \{[C_1 p]; C_1 p \varepsilon \mathfrak{A}_1\} \varepsilon \mathfrak{Q}$;

(5.9) if $\cup \mathfrak{A}_2 \varepsilon C_2\mathfrak{B}$, then $\cup \{[C_2 p]; C_2 p \varepsilon \mathfrak{A}_2\} \varepsilon \mathfrak{Q}$;

(5.10) $\pi_0 \varepsilon \mathfrak{Q}$.

Let $\mathfrak{Q}$ be the sub-Boolean-algebra of $\mathfrak{Q}'$ generated by these prescribed elements.

This ensures that $\mathfrak{Q}$ is closed under the Boolean operations. The proof that it is closed under the other operations follows.

Any $\alpha \varepsilon \mathfrak{Q}$ may be written as the join of the $\alpha \cap \chi_n$, $n = 1, 2, 3, 4, 5$. Also $\alpha \cap \chi_n \varepsilon \mathfrak{Q}$, and, by P1, $(\alpha \cap \chi_n)_x \varepsilon \mathfrak{Q}$ for all n implies $\alpha_x \varepsilon \mathfrak{Q}$. So the next task is to prove, for $\alpha \subseteq \chi_n$, that $\alpha \varepsilon \mathfrak{Q}$ implies $\alpha_x \varepsilon \mathfrak{Q}$.

For $n = 5$, that is, $\alpha = \pi_0$, this is automatic.

For $n = 1, 3, 4$, it is almost as obvious. Take for example $\alpha \subseteq \chi_1$, $\alpha \varepsilon \mathfrak{Q}$. It is easy to see that α is of the form (5.6): $\alpha = \cup \{[p]; p \subseteq a, a \varepsilon \mathfrak{B}\}$. Then $\alpha_x = \cup \{[C_1 p]; C_1 p \subseteq C_1 a\}$; since $C_1 a \varepsilon C_1\mathfrak{B}$, this is of the form (5.8), hence in $\mathfrak{Q}$.

Finally, consider $\alpha \subseteq \chi_2$, $\alpha \varepsilon \mathfrak{Q}$. Call an element of the form (5.7) a "$\mathfrak{Q}$-rectangle." Now α is a combination of a finite number of $\mathfrak{Q}$-rectangles, together with elements disjoint from χ_2, under Boolean operations. Note that, for any $\mathfrak{Q}$-rectangles β and γ, $\beta \cap \gamma$ and $\beta' \cap \chi_2$ are joins of $\mathfrak{Q}$-rectangles. Therefore α is a join of $\mathfrak{Q}$-rectangles; and therefore, by P1, only the case where α is a single $\mathfrak{Q}$-rectangle needs to be considered. But this case is clear.

Thus $\mathfrak{Q}$ is closed under x-projection; similarly for y-projection. There remains the operation "$\square$."

Assume $\alpha = \cup\{[C_1 p]; C_1 p \subseteq a \varepsilon C_1\mathfrak{B}\}$, $\beta = \cup\{[C_2 p]; C_2 p \subseteq b \varepsilon C_2\mathfrak{B}\}$; so that $\alpha, \beta \varepsilon \mathfrak{Q}$. In order for a point π of $\mathfrak{Q}'$ to be below $\alpha \square \beta$, it must

satisfy $\pi_x \subseteq \alpha$ and $\pi_y \subseteq \beta$. This is equivalent to saying that *either* $\pi = [p]$, with $p \subseteq a \cap b$, *or* $\pi = [C_1 p, C_2 q]$, with $C_1 p \subseteq a$, $C_2 q \subseteq b$. The join of all π of the first kind is an element of the form (5.6), because a and b, hence $a \cap b$, are in $\mathfrak{B}$; the join of all π of the second kind is an element of the form (5.7). Therefore $\alpha \square \beta$, the join of all π of either kind, is in $\mathfrak{Q}$.

This was not quite general enough, it assumed $\pi_0 \not\subseteq \alpha$, $\pi_0 \not\subseteq \beta$. However the remaining cases involve no difficulty, and one concludes that $\mathfrak{Q}$ is closed under the operation "$\square$."

It is now easy to see that $\mathfrak{Q}$ is a projective algebra.

Part III. The first two parts of the proof were preliminary: the supplying of a projective algebra $\mathfrak{Q}$ from which an isomorphic image of $\mathfrak{B}$ could be obtained by the construction of Theorem 4.1. There remains the proof proper: showing that this can in fact be done using the $\mathfrak{Q}$ prescribed.

Define the function Φ on $\mathfrak{B}'$ by $\Phi a = \cup \{[p]; p \subseteq a, p \text{ a point}\}$. Then $\Phi\mathfrak{B}'$ consists of those elements of $\mathfrak{Q}'$ below ΦI, which is the element called χ_1 above. Any $\alpha \in \Phi\mathfrak{B}'$ is in $\Phi\mathfrak{B}$ if and only if it is in $\mathfrak{Q}$; and Φ is a Boolean isomorphism of $\mathfrak{B}$ on $\Phi\mathfrak{B}$ (in fact of $\mathfrak{B}'$ on $\Phi\mathfrak{B}'$), if complementation in $\Phi\mathfrak{B}'$ is understood as relative to ΦI.

So much is obvious. Now by Theorem 4.1, $\bar{C}_1$ and $\bar{C}_2$ defined as follows are $S5$ operators on $\Phi\mathfrak{B}$: $\bar{C}_1\alpha = (\alpha_x \square i_y) \cap \Phi I$, $\bar{C}_2\alpha = (\iota_x \square \alpha_y) \cap \Phi I$. It must be shown that $\Phi C_1 a = \bar{C}_1 \Phi a$, $\Phi C_2 a = \bar{C}_2 \Phi a$.

Assume $[p] \subseteq \bar{C}_1\alpha$, $\alpha = \Phi a$, $a \in \mathfrak{B}$. Since $[p] \subseteq \alpha_x \square \iota_y$, one has $[p]_x \subseteq \alpha_x$; and every point below α_x is $[C_1 q]$ for some $[q] \subseteq \alpha$, that is, for $q \subseteq a$. Hence, for such a q, $p \subseteq C_1 p = C_1 q \subseteq C_1 a$. This proves $\Phi C_1 a \supseteq \bar{C}_1 \Phi a$.

Let p be a point of $\mathfrak{B}'$ below $C_1 a$, $a \in \mathfrak{B}$; then $[p] \subseteq \Phi C_1 a$. By the proof of Theorem 2.2, there is some point q below a such that $p \subseteq C_1 q$. For such a q, $[p]_x = [C_1 p] = [C_1 q] = [q]_x \subseteq (\Phi a)_x$. Thus p, being a point below ΦI whose x-projection is below $(\Phi a)_x$, is below $(\Phi a \square \iota_y) \cap \Phi I = \bar{C}_1 \Phi a$. This proves $\Phi C_1 a \subseteq \bar{C}_1 \Phi a$.

An identical argument relating C_2 with $\bar{C}_2$ completes the proof of the theorem.

It should be remarked that in case e of Theorem 4.1 is i, then for any $a \neq O$, $C_2 C_1 a = C_2(a_x \square i_y) = i_x \square i_y = i$ (see [7, C10]). But there exist $\mathfrak{B}$ such that $C_2 C_1 a$ assumes other values than O and I. Therefore if Theorem 4.1 had asserted only the special case $e = i$, the converse Theorem 4.2 would have been false.

6. Relation to the first-order functional calculus. This relation is most easily expressed by dealing with Mautner's Boolean tensors [13] instead of with the logical theory itself. The logical coordinate system of Mautner's theory will be left fixed,[7] so that a Boolean r-tensor may be treated as simply a function on the set $S \times S \times \cdots \times S$ (r factors) to the two-element Boolean algebra.

The Boolean r-tensors $f = f(x_1, \cdots, x_r)$, $x_k \varepsilon S$, form a system $\mathfrak{B}_r$ closed under the Boolean operations. However $\mathfrak{B}_r$ is not closed under the following operations:

(6.1) outer sum formation

$$f(x_1, \cdots, x_r) \cup g(y_1, \cdots, y_r) = h(x_1, \cdots, x_r, y_1, \cdots, y_r) \varepsilon \mathfrak{B}_{2r};$$

(6.2) outer product formation, dual to (6.1);

(6.3) setting arguments equal

$$f(\cdots, x_i, \cdots, x_j, \cdots) \to f(\cdots, x_i, \cdots, x_i, \cdots);$$

(6.4) sum contraction (i.e., existential quantification)[8]

$$\cup \{f(x_1, \cdots, x_r); x_k \varepsilon S\} = g(x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_r) \varepsilon \mathfrak{B}_{r-1};$$

(6.5) product contraction, dual to (6.4).

All of (6.1)-(6.5) correspond to admissible processes in forming formulas of the functional calculus.[9]

If all of (6.1)-(6.5) are forbidden, $\mathfrak{B}_r$ is merely a Boolean algebra. This corresponds to the situation of the "alternate universes" of Section 3 before $S5$ operators were introduced. Corresponding to the introduction of $S5$ operators there, a *limited class* of Boolean tensor operators will be specified which *will* take $\mathfrak{B}_r$ into itself. (Note that (6.1)-(6.3) will still be forbidden, and will have no analogue.) Namely, for any $y \varepsilon S$ let

(6.6) $$\cup \{f(x_1, \cdots, x_r); x_k \varepsilon S\} = (C_k f)(x_1, \cdots, x_{k-1}, y, x_{k+1}, \cdots, x_r).$$

Clearly $C_k f$ is in $\mathfrak{B}_r$, but its k-th argument is without effect. Conversely,

[7] It will be clear that Theorem 6.1 *has invariant significance* in Mautner's sense and Theorem 6.2 does not. Note that my notation for Boolean tensors in this section differs from Mautner's.

[8] The notation in the left-hand member implies that the k-th variable is bound by the summation.

[9] It is unconventional to include setting arguments equal (6.3) as a process analogous to the others; that is, "$F(x, x)$" is not usually regarded as having been formed from "$F(x, y)$." But cf. [9; Kap. 5a].

a function in $\mathfrak{B}_r$ whose value is independent of its k-th argument is of the form $C_k f$. This characterization of $C_k\mathfrak{B}_r$ makes it clear that it is a complete sub-Boolean-algebra of $\mathfrak{B}_r$. Also C_k can be seen to take any $f \in \mathfrak{B}_r$ to the meet of those elements of $C_k\mathfrak{B}_r$ above f. Therefore (Theorem 2.1) C_k is an $S5$ operator.

Furthermore, if ξ is any set $\{k_1, \cdots, k_t\}$ of the indices $1, \cdots, r$, then the operator

$$C_\xi = C_{k_1} C_{k_2} \cdots C_{k_t} \tag{6.7}$$

clearly depends only on ξ (not on the order of the C_k) and is an $S5$ operator.

Some further properties of the C_ξ may be pointed out. It is natural to order them as functions into $\mathfrak{B}_r$, in fact to consider them as a subset of the lattice of all $S5$ operators on $\mathfrak{B}_r$ (Section 2 and [1]). Now if one defines an operator corresponding to the null set of indices by

$$C_0 f = f, \tag{6.8}$$

then one concludes that $C_{\xi \cup \eta} = C_\xi \cup C_\eta$. It is not hard to show also that $C_{\xi \cap \eta} = C_\xi \cap C_\eta$. (The proof, based on Theorem 2.1, is given in [5].) Finally, if $\xi = \{1, \cdots, r\}$, then $C_\xi f(x_1, \cdots, x_r) = I$ for all $x_1, \cdots, x_r$, provided there is any set of arguments for which $f(x_1, \cdots, x_r) = I$.

Collecting these facts gives the following.

THEOREM 6.1.[7] *The operators C_ξ on $\mathfrak{B}_r$ defined by* (6.6)-(6.8) *are $S5$ operators. They form a sublattice of the lattice of all $S5$ operators* (containing its null and unit elements), *and $C_\xi \to \xi$ is a lattice-isomorphism onto the Boolean algebra of all sets of indices* $1, \cdots, r$.

Thus Boolean algebras with $S5$ operators generalize Boolean tensors only if the operations allowed on the latter are restricted; they provide a Boolean model for the functional calculus only in a limited sense. Quantification is included (notice that the dual of (6.6) is $(C_k f')'$) but (6.1)-(6.3) are not.

Projective algebras are evidently in the same situation. In fact, identify every $f \in \mathfrak{B}_2$ with the set of all points of $S \times S$ which it maps into I. Fix a point z_0 of S. Define[10] f_x as the set-intersection of $\{z_0\} \times S$ with $C_1 f$ (as defined by (6.6)); similarly $f_y = (S \times \{z_0\}) \cap C_2 f$. Also define, for subsets T and U of S.

$$(\{z_0\} \times T) \square (U \times \{z_0\}) = U \times T \subseteq S \times S.$$

---

[10] [7, C28].

THEOREM 6.2.[7] *With respect to the distinguished point* $\{z_0\} \times \{z_0\}$ *and the operations just defined,* $\mathfrak{B}_2$ *becomes the projective algebra of all subsets of the direct product* $S \times S$. *Every projective algebra of all subsets of a direct product of a set with itself is obtained in this way.*

This is clear without giving details. For the first sentence, compare [7]. For the second sentence, simply construct $\mathfrak{B}_2$ over the set in question.

ANN ARBOR, MICHIGAN.

BIBLIOGRAPHY.

[1] G. Birkhoff, *Lattice theory*, American Mathematical Society Colloquium Publications, vol. 25 (1948), 2nd ed., N. Y.

[2] R. Carnap, "Modalities and quantification," *Journal of Symbolic Logic*, vol. 11 (1946), pp. 33-64.

[3] ———, *Meaning and necessity*, Chicago, 1947.

[4] L. H. Chin and A. Tarski, "Remarks on projective algebras," *Bulletin of the American Mathematical Society*, Abstract 54-1-90 (1948).

[5] C. Davis, *Lattices and modal operators*, Harvard Doctoral Thesis, 1950.

[6] A. F. Emch, "Implication and deducibility," *Journal of Symbolic Logic*, vol. 1 (1936), pp. 26-35; also *ibid.*, p. 58.

[7] C. J. Everett and S. Ulam, "Projective algebra I," *American Journal of Mathematics*, vol. 68 (1946), pp. 77-88.

[8] D. Hilbert and W. Ackermann, *Grundzüge der theoretischen Logik*, 2nd ed., Berlin, 1938.

[9] D. Hilbert and P. Bernays, *Grundlagen der Mathematik. I Band*, Berlin, 1934.

[10] B. Jónsson and A. Tarski, "Boolean algebras with operators," *Bulletin of the American Mathematical Society*, Abstract 54-1-88 (1948).

[11] C. I. Lewis, "Emch's calculus and strict implication," *Journal of Symbolic Logic*, vol. 1 (1936), pp. 77-86.

[12] C. I. Lewis and C. H. Langford, *Symbolic logic*, New York, 1932.

[13] F. I. Mautner, "An extension of Klein's Erlanger Program: logic as invariant theory," *American Journal of Mathematics*, vol. 68 (1946), pp. 345-384.

[14] J. C. C. McKinsey, "On the syntactical construction of systems of modal logic," *Journal of Symbolic Logic*, vol. 10 (1945), pp. 83-94.

[15] J. C. C. McKinsey, "On the representation of projective algebras," *American Journal of Mathematics*, vol. 70 (1948), pp. 375-384.

[16] J. C. C. McKinsey and A. Tarski, "The algebra of topology," *Annals of Mathematics*, vol. 45 (1944), pp. 141-191.

[17] J. C. C. McKinsey and A. Tarski, "Some theorems about the sentential calculi of Lewis and Heyting," *Journal of Symbolic Logic*, vol. 13 (1948), pp. 1-15.

[18] O. Ore, "Theory of equivalence relations," *Duke Mathematical Journal*, vol. 9 (1942), pp. 573-627.

[19] O. Ore, "Some studies of closure relations," *Duke Mathematical Journal*, vol. 10 (1943), pp. 761-785.

[20] Tang Tsao-Chen, "Algebraic postulates and a geometrical interpretation for the Lewis calculus of strict implication," *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 737-744.

[21] P. G. J. Vredenduin, "A system of strict implication," *Journal of Symbolic Logic*, vol. 4 (1939), pp. 73-75.

[22] M. Wajsberg, "Ein erweiterter Klassenkalkül," *Monatshefte für Mathematik und Physik*, vol. 40 (1932), pp. 113-126; also *ibid.*, vol. 42 (1935), p. 242.

[23] H. Weyl, "The ghost of modality," in *Philosophical essays in memory of Edmund Husserl*, ed. Marvin Farber, Cambridge (Mass.), 1940, pp. 278-303.

COHOMOLOGY CLASSES OF FINITE TYPE AND FINITE DIMENSIONAL KERNELS FOR LIE ALGEBRAS.*

By G. HOCHSCHILD.

Introduction. Our main concern is with the problem of finite dimensionality of the kernels in the interpretation of the 3-dimensional cohomology groups of a Lie algebra L as groups of equivalence classes of L-kernels (as in [4]), in analogy with the theory of group kernels of Eilenberg-MacLane, [1]. This problem is connected with the theorem on the existence of a finite dimensional faithful representation of a finite dimensional Lie algebra and with the related problem of the effaceability of cohomology classes in finite dimensional modules. In fact, the solution of our problem depends on some of the methods and results of several investigations in these directions.

In order to give a coherent and intelligible account of these and certain refinements of them which we need for our purpose, we take up some of their essential details in a more specifically adapted exposition. In particular (in Section 1) we give a new exposition of the relevant features of the use of the differential calculus of E. Cartan which was basic in [2] and [9] ([9] is not available in print), cutting out much of the computational detail which encumbered the original presentation. In Section 2, we give a slight refinement of a result on module enlargements due to Zassenhaus, [10], which is vital for our purpose. In Section 3, this is combined with the results of Section 1 to give a completion of the result of [8] and [9] on the effaceability of cohomology classes. This will absorb Iwasawa's result, [6], on the effaceability of 2-dimensional cohomology classes in characteristic 0. We also give a simplification of Iwasawa's proof for the case of characteristic p by using Jacobson's technique of [7]. In Section 4, we apply these results to the kernel problem for which we obtain a complete solution, in the case of characteristic 0.

1. Cohomology and differential forms. Let L be a finite dimensional Lie algebra over a field F. Denote by A the exterior algebra built over the space dual to L. A is a graded algebra over F, with the homogeneous com-

---

* Received March 22, 1954.

ponents A^m such that $A^m = (0)$, for $m < 0$; $A^0 = F$; $A^1 =$ the dual space of L; and, for $m \geqq 1$, $A^m =$ the subspace of A which is spanned by the (exterior) products with m factors from A^1. The (exterior) multiplication in A is anticommutative in the sense that, for $a \in A^p$ and $b \in A^q$, $ba = (-1)^{pq}ab$. Without dwelling on the well known details of the definition of A, we recall merely that A^m may be identified with the space of all m-linear alternating functions from L to F. In particular, $A^m = (0)$ if m exceeds the dimension of L, and otherwise the ordered products of m elements of an ordered basis for A^1 constitute a basis for A^m over F. The coboundary operator δ of A is a homogeneous antiderivation of degree 1, which means that $\delta(A^m) \subset A^{m+1}$ and $\delta(ab) = \delta(a)b + (-1)^p a\delta(b)$, if $a \in A^p$. Furthermore, $\delta(F) = (0)$, and $\delta^2 = 0$. Clearly, such a map δ will be uniquely determined (assuming its existence) as soon as its values on A^1 have been prescribed. This is done most naturally by using the following further maps.[1] For $u \in L$, define the map c_u of A^1 into F by $c_u(a) = a(u)$. Then c_u can be extended in one and only one way to a homogeneous antiderivation of degree -1 of A, which we still denote by c_u. It follows then at once by an induction on the degree that $c_u^2 = 0$. On the other hand, there is a unique homogeneous derivation t_u of degree 0 on A which maps F into (0) and, on A^1, is given by $t_u(a) = a_u \in A^1$, where $a_u(v) = a([v, u])$. One verifies easily that $t_u t_v - t_v t_u = t_{[u,v]}$, so that A carries the structure of an L-module. This verification is again based on an induction on the degree, noting that the two maps whose equality is to be established are both derivations of degree 0. Similarly, one observes that $t_u c_v - c_v t_u$ and $c_{[u,v]}$ are antiderivations of degree -1 which coincide on F and on A^1, whence it follows that $t_u c_v - c_v t_u = c_{[u,v]}$. Now the conditions $c_u(\delta a) = t_u(a)$, for all $u \in L$ and all $a \in A^1$ define a map $\delta: A^1 \to A^2$ uniquely. There is unique extension of this map to a homogeneous antiderivation of degree 1 of A which is 0 on F. This is the coboundary map δ of A, and one verifies easily that $c_u\delta + \delta c_u = t_u$, for all $u \in L$, and that $\delta^2 = 0$. This verification is again inductive, using the fact that $c_u\delta + \delta c_u$ and δ^2 are derivations of A.

Now let V be an L-module. We form the tensor product $V \otimes A$ with respect to the base field F. We may regard this as a right A-module in the natural fashion, and we shall "extend" the above maps c_u, t_u and δ over $V \otimes A$ in such a way that their inter-relations are preserved, as well as the properties $c_u^2 = 0$, $\delta^2 = 0$. We shall define c_u and t_u on $V \otimes A$ such that, for

[1] These maps and, more generally, the technique of derivations and antiderivations in graded rings which we employ here have been used quite frequently in recent years, notably by H. Cartan, C. Chevalley, and J. L. Koszul. A certain familarity with their manipulation is presupposed here.

$v \in V$ and $a \in A$, $c_u(v \otimes a) = v \otimes c_u(a)$, and $t_u(v \otimes a) = (u \cdot v) \otimes a + v \otimes t_u(a)$. The coboundary map δ is extended so that $\delta(v \otimes a) = (\delta v)a + v \otimes \delta a$, where δv is defined as an element of $V \otimes A^1$ by the relations $c_u(\delta v) = u \cdot v$, for all $u \in L$. In particular, the pair $(V \otimes A, \delta)$ constitutes a cochain complex whose cohomology groups are the ordinary cohomology groups $H^n(L, V)$ for L in V.

Next we single out certain facts concerning rings of power series whose consideration is basic for our proof of the effaceability theorem. Let P denote the ring of all integral formal power series with coefficients in the field F of characteristic 0 in n variables $x_1, \cdots, x_n$. We denote by D_i the partial differentiation with respect to x_i in P. Correspondingly, we denote the formal quadrature with respect to x_i by D_i^{-1} (the "constant of integration" is taken to be 0). If U is an F-linear subspace of P, the *calculus closure* of U in P is defined to be the subspace of P which is spanned by all those elements of P that can be obtained from elements of U by applying a finite number of operators D_i and D_i^{-1}. The differentiations of P (over F) constitute a free P-module with the D_i as a free basis. An element $p \in P$ gives a P-linear map $D \to D(p)$ of this module into P which we denote by dp; this is the differential of p. The differentials generate (over P) a free P-module which is dual to the module of the differentiations of P; $dx_1, \cdots, dx_n$ being a free basis dual to the free basis $D_1, \cdots, D_n$ of the module of differentiations. The algebra B of the differential forms on P is the exterior algebra built (relative to P) over the module generated by the differentials, which is therefore B^1. We have $B^0 = P$, and d maps B^0 (F-linearly) into B^1. This map d can be extended uniquely to an antiderivation of degree 1 of B, still denoted d, such that $d^2 = 0$. As with every graded ring, we denote by B^+ the sum of all homogeneous components of positive degree of B. Then B^+ is an ideal of B, and B is the direct sum of B^0 and B^+. The most important fact for us is the following well known result, a proof of which can be found, for instance, in Section 2 of [2].

LEMMA 1. *Let b be an element of B^+ such that $db = 0$. Let U_b be the subspace of P which is spanned by the coefficients of b when expressed as a P-linear combination of the ordered products of the dx_i. Then there exists an element $c \in B$ such that $dc = b$ and U_c lies in the calculus closure of U_b.*

Now let L be a solvable Lie algebra of finite dimension n over an algebraically closed field F of characteristic 0. Let K be another Lie algebra over F which contains L as an ideal. Then, since simple L-modules must be 1-dimensional (as follows from the fact that L is solvable and F

algebraically closed), we can find a sequence $(0) = L_n \subset L_{n-1} \subset \cdots \subset L_r = [K, L] \subset L_{r-1} \subset \cdots \subset L_0 = L$, where each L_i is an ideal in L and has dimension $n - i$. We select a basis $u_1, \cdots, u_n$ for L over F such that each L_i is spanned by $u_{i+1}, \cdots, u_n$. Then, if we write $[u_i, u_j] = \sum_{k=1}^{n} c_{ijk} u_k$, with $c_{ijk} \in F$, we have $c_{ijk} = 0$, unless $k \geqq i$, $k \geqq j$, and $k > r$.

Applying Lie's theorem to the solvable Lie algebra spanned by L and a single element x of K, we conclude that $[x, L]$ is nilpotent on every finite dimensional K-module. Hence $[[x, L], L_{j-1}] \subset L_j$ $(j = 1, \cdots, n)$ for every $x \in K$, whence $[[K, L], L_{j-1}] \subset L_j$. We have therefore $c_{ijj} = 0$ if $i > r$.

Next we consider a finite dimensional K-module V. Regarding V as an L-module, we can (with the same justification as was used above) find a basis $v_1, \cdots, v_m$ for V which is such that, for each k, $L \cdot v_k$ is contained in the space spanned by $v_k, \cdots, v_m$. It follows then as above that $[K, L] \cdot v_k$ is contained in the space spanned by $v_{k+1}, \cdots, v_m$. If we write $u_i \cdot v_j = \sum_{k=1}^{m} e_{ijk} v_k$, with e_{ijk} in F, we have therefore $e_{ijk} = 0$, unless $k \geqq j$; and, for $i > r$, $e_{ijj} = 0$. These facts concerning the coefficients e_{ijk} and c_{ijk} will sometimes be used implicitly in what follows.

Certain elements of our power series ring P will play a distinguished role below. These are the following exponential series: $c_i = \exp(\sum_{p=1}^{r} c_{pii} x_p)$, and $e_j = \exp(\sum_{p=1}^{r} e_{pjj} x_p)$. We denote by Q the subring of P which is generated by F, the x_k $(k = 1, \cdots, n)$, the c_i and their inverses $(i = 1, \cdots, n)$, and the e_j and their inverses $(j = 1, \cdots, m)$. Correspondingly, we replace B by the subring C of B which is generated by Q and the differentials dx_k. Evidently, Q coincides with its calculus closure. From Lemma 1, it follows that if b is an element of C^+ such that $db = 0$ there is an element c in C such that $dc = b$.

We shall construct an isomorphism of the complex (A, δ) into the complex (C, d). Let $a_1, \cdots, a_n$ be the basis of A^1 which is dual to the basis $u_1, \cdots, u_n$ of L, so that $a_i(u_j) = \delta_{ij}$ (Kronecker symbol). Then the definition of δ shows that $\delta a_i = -\sum_{j<k} c_{jki} a_j a_k$. If an isomorphism of (A, δ) into (C, d) leaves the elements of F fixed and sends each a_i into b_i, we must therefore have $db_i = -\sum_{j<k} c_{jki} b_j b_k$. We can write these conditions in the form $db_i + \sum_{j=1}^{i-1} c_{jii} b_j b_i = -\sum_{j<k<i} c_{jki} b_j b_k$. For $i \leqq r$, these conditions become simply $db_i = 0$. Accordingly, we shall define $b_i = dx_i$, for $i \leqq r$. Next we observe that $c_{jii} = 0$ if $j \geqq i$, and also if $j > r$. Hence, if $i > r$, our conditions

on the b_i may be written $db_i + d(\sum_{p=1}^{r} c_{pii}x_p)b_i = -\sum_{j<k<i} c_{jki}b_jb_k$. These conditions are equivalent to the conditions $d(c_ib_i) = -c_i\sum_{j<k<i} c_{jki}b_jb_k$.

Now suppose that $i > r$, and that we have already found elements $b_1, \cdots, b_{i-1}$ in C^1 which satisfy the above conditions and which are furthermore of the form $b_s = c_s^{-1}dx_s + \sum_{t=1}^{s-1} q_{st}dx_t$, where the q_{st} are elements of Q which do not involve $x_s, \cdots, x_n$. Write $p_i = \sum_{j<k<i} c_{jki}b_jb_k$, and, correspondingly, $P_i = \sum_{j<k<i} c_{jki}a_ja_k$. Similarly, set $r_i = \sum_{p=1}^{r} c_{pii}dx_p$, and $R_i = \sum_{p=1}^{r} c_{pii}a_p$. We have $\delta a_i + R_ia_i + P_i = 0$. Hence, applying δ, $(\delta R_i)a_i - R_i\delta a_i + \delta P_i = 0$. Substituting $R_ia_i + P_i$ for $-\delta a_i$ and observing that $R_i^2 = 0$ (for $R_i \varepsilon A^1$), we obtain $(\delta R_i)a_i + R_iP_i + \delta P_i = 0$. Since $R_iP_i + \delta P_i$ involves only the a_j's with $j < i$, this implies that we must have $R_iP_i + \delta P_i = 0$. Evidently, it follows from this that we have also $r_ip_i + dp_i = 0$. But this means that $d(c_ip_i) = 0$. By Lemma 1, we can find elements q_i in Q such that $dq_i = -c_ip_i$, and furthermore such that $q_i = c_i\sum_{t=1}^{i-1} q_{it}dx_t$, where the q_{it} are as in our inductive assumption. (Note that we have $i > r$, and that $x_i, \cdots, x_n$ do not enter in this consideration.) Now we define $b_i = c_i^{-1}(dx_i + q_i)$. Thus we can find $b_1, \cdots, b_n$ in C^1 which satisfy our above conditions and which are furthermore of the form described in our inductive assumption. From this form of the b_i's, it is clear that the dx_i's can be expressed as Q-linear combinations of the b_i's, whence the b_i's constitute a free basis for C^1 over Q. The map which leaves the elements of F fixed and sends each a_i into b_i is therefore an F-linear isomorphism of $A^0 + A^1$ into $C^0 + C^1$ and extends uniquely to a ring isomorphism ϕ of A into C. By the construction of the b_i's, we have $d\phi = \phi\delta$ on $A^0 + A^1$, whence it follows that $d\phi = \phi\delta$ on all of A. Thus ϕ is an isomorphism of the complex (A, δ) into the complex (C, d). Now we define, for each $u \varepsilon L$, an antiderivation C_u of degree -1 on C such that $\phi c_u = C_u\phi$. In fact, we define $C_u(\phi(a)) = c_u(a) = a(u)$, for $a \varepsilon A^1$. Since the b_i's constitute a free basis for C^1 over Q, this map C_u extends uniquely to an antiderivation of degree -1 of C which clearly has the desired property. Now we define the derivations T_u of degree 0 of C by setting $T_u = C_ud + dC_u$. We have then $T_u\phi = \phi t_u$. Hence we find also $(T_uC_v - C_vT_u)\phi = C_{[u,v]}\phi$, for all u and v in L. This shows that the antiderivation of degree -1, $T_uC_v - C_vT_u - C_{[u,v]}$, vanishes on C^1, whence we may conclude that $T_uC_v - C_vT_u = C_{[u,v]}$, on all of C. Observing that $T_ud = dT_u$, we derive from this, by writing $T_v = C_vd + dC_v$, that $T_uT_v - T_vT_u = T_{[u,v]}$.

In particular, this shows that Q becomes an L-module with the operators T_u. It is easily seen from our definitions that, for $i \leqq r$, $T_{u_j}(x_i) = \delta_{ij}$. Hence T_u maps each exponential of a linear form in $x_1, \cdots, x_r$ into an F-multiple of itself, which is zero whenever $u \in [K, L]$. Furthermore, it follows inductively from the form of the b_i's that, for any i, and any $u \in L$, $T_u(x_i)$ does not involve $x_i, \cdots, x_n$. From these facts, it is not difficult to prove the following critical result.

LEMMA 2. *Every finite dimensional subspace (relative to F) of Q is contained in a finite dimensional L-submodule of Q. Furthermore, $[K, L]$ is nilpotent on every finite dimensional L-submodule of Q.*

Proof. Let $M > 1$ be an upper bound for the total degrees in the x_k's of the elements $T_{u_j}(x_i)$. In counting degrees, the exponentials are ignored, i. e., they are regarded as elements of degree 0. The weight $w(q)$ of a monomial $q = {x_1}^{m_1} \cdots {x_n}^{m_n}$ is defined as $w(q) = \sum_i m_i M^{2i}$. If q is an arbitrary element of Q, $w(q)$ is defined as the maximum of the weights of the monomials occurring in q. If q is an ordinary polynomial in $F[x_1, \cdots, x_n]$ the weight of $T_u(q) = \sum_i D_i(q) T_u(x_i)$ is at most equal to the maximum of the n numbers $w(q) - M^{2i} + MM^{2(i-1)}$, whence $w(T_u(q)) < w(q)$, for all $q \in F[x_1, \cdots, x_n]$. If q is a polynomial and e an exponential we have $T_u(eq) = \alpha eq + eT_u(q)$, where $\alpha \in F$; and $\alpha = 0$ if $u \in [K, L]$. It is now clear how one can complete the proof of Lemma 2 by a simple accounting of the weights.

Now we form the tensor product $V \otimes C$ relative to F and make a complex $(V \otimes C, d)$ by extending d over $V \otimes C$ in the same way as we extended δ in our construction of the complex $(V \otimes A, \delta)$. Our isomorphism ϕ of (A, δ) into (C, d) evidently yields an isomorphism of $(V \otimes A, \delta)$ into $(V \otimes C, d)$ which sends $v \otimes a$ into $v \otimes \phi(a)$. On the other hand, we shall prove that the complex $(V \otimes C, d)$ is isomorphic with the complex (C_m, d), where C_m is the direct sum of m copies of C, and d is applied coordinatewise on C_m. We shall treat C_m as a C-module in the natural fashion, and our isomorphism will be C-linear. In $V \otimes C$, we have $dv_j = \sum_i u_i \cdot v_j \otimes b_i = \sum_{i,k} e_{ijk} v_k \otimes b_i$. Hence a C-linear isomorphism of $(V \otimes C, d)$ into (C_m, d) must map the v_j into elements w_j of ${C_m}^0 = Q_m$ which satisfy the relations $dw_j = \sum_{i,k} e_{ijk} w_k b_i$. These relations may be written in the form $dw_j - w_j \sum_{p=1}^{r} e_{pjj} dx_p = \sum_{i=1}^{n} \left(\sum_{k>j} e_{ijk} w_k \right) b_i$, or, equivalently, $d(w_j e_j^{-1}) = \sum_{i=1}^{n} \left(\sum_{k>j} e_{ijk} w_k \right) b_i e_j^{-1}$. Let y_j denote the element

of Q_m whose j-th component is equal to 1, while all its other components are equal to 0. For $j = m$, our above condition becomes simply $d(w_m e_m^{-1}) = 0$. Accordingly, we define $w_m = y_m e_m$. Now suppose that we have already found elements $w_m, w_{m-1}, \cdots, w_{j+1}$ such that the above relations are satisfied, and furthermore such that each w_k is of the form $w_k = y_k e_k + f_k$, where $f_k \in y_{k+1}Q + y_{k+2}Q + \cdots + y_m Q$. Put $p_j = \sum_i (\sum_{k>j} e_{ijk} v_k) \otimes b_i$, $P_j = \sum_i (\sum_{k>j} e_{ijk} w_k) b_i$, and $r_j = \sum_p e_{pjj} dx_p$. Then we have $dv_j = v_j \otimes r_j + p_j$. Applying d, we obtain $0 = d(v_j) r_j + dp_j$. Substituting $v_j \otimes r_j + p_j$ for dv_j, we get $p_j r_j + dp_j = 0$. From this it evidently follows that we have also $P_j r_j + dP_j = 0$, whence $d(P_j e_j^{-1}) = 0$. By Lemma 1, there is an element h_j in $y_{j+1}Q + y_{j+2}Q + \cdots + y_m Q$ such that $dh_j = P_j e_j^{-1}$. We set $w_j = (y_j + h_j) e_j$. Hence we can successively find $w_m, \cdots, w_1$ which satisfy our above relations and are of the form described above. From the form of these elements, it is clear that they constitute a free basis for C_m over C. The C-linear map which sends each v_j into w_j is now seen to be an isomorphism of the complex $(V \otimes C, d)$ onto the complex (C_m, d). This gives the following main result of this section.[2]

LEMMA 3. *Let K be a Lie algebra over a field F of characteristic 0, and let L be a solvable finite dimensional ideal of K. Let V be a finite dimensional K-module. Then there exists a finite dimensional L-module U on which $[K, L]$ is nilpotent, which contains V as an L-submodule, and which is such that the canonical homomorphism $H(L, V) \to H(L, U)$ annihilates every cohomology class of positive dimension.*

Proof. It is easily seen that there is no loss of generality in supposing that F is algebraically closed. Then we can use the above imbedding of $(V \otimes A, \delta)$ in (C_m, d). Clearly, Lemma 1 extends to C_m. Hence every cohomology class for L in V which is of positive dimension is represented in C_m by an element of the form dc. The coefficients in Q_m of c when expressed in terms of the b_i's, together with the elements of V (identified with a subspace of Q_m), span a finite dimensional subspace of Q_m which, by Lemma 2, is contained in a finite dimensional L-submodule of Q_m. Since the vector space $H(L, V)$ is finite dimensional, it is now clear that the L-module U of Lemma 3 can be found as an L-submodule of Q_m.

---

[2] The result of Lemma 3, except for the fact that U can be chosen so that $[K, L]$ is nilpotent on U, is proved by a very different method in [8]. The proof given here is essentially that of [9].

2. Zassenhaus' module enlargement. We shall require a slight refinement of a result of Zassenhaus on enlargements of modules. In proving it, we follow Zassenhaus quite closely.

Let S be a finite dimensional Lie algebra over an arbitrary field F, and let R_S denote the universal enveloping algebra of S. R_S is generated by 1 and the elements of S. We define an increasing filtration of R_S by subspaces R_m such that $R_pR_q \subset R_{p+q}$, as follows: for $m < 0$, $R_m = (0)$; $R_0 = F$; for $m \geqq 1$, R_m is the subspace of R_S which is spanned by all products (in R_S) of no more than m elements of S. Then the associated graded ring $\sum_m R_m/R_{m-1}$ is isomorphic with the symmetric algebra built over S, i. e., with a polynomial algebra in n variables, where n is the dimension of S. From this one verifies easily that the maximal condition holds for the ideals (left, right, or two sided) of R_S. On the other hand, one shows easily that, if R is any algebra over F and A and B are ideals of finite codimension in R (i. e., R/A and R/B are finite dimensional) and if, furthermore, B is finitely generated as an R-module, AB is also of finite codimension in R. Hence one concludes that, if A and B are ideals of finite codimension in R_S, so is AB.

Now consider a finite dimensional Lie algebra K which is of the form $K = S + L$, where S is a subalgebra and L an ideal in K, and $S \cap L = (0)$. Then we have $R_S \subset R_K$, $R_L \subset R_K$, and $R_K = R_SR_L = R_LR_S$. From now on, the word ideal will mean two sided ideal. If C is an ideal of R_S, we denote by C' the set of all elements $c \in C$ for which the commutator space $[L, c]$ is contained in $C[S, L]$. It is not difficult to verify that C' is still an ideal of R_S. Furthermore, if C is of finite codimension in R_S, so is C'. In fact, let $z_1, \cdots, z_q$ be a basis for L which contains a basis $z_1, \cdots, z_p$ for $[S, L]$. Now observe that $1 \in R_S'$, whence $R_S' = R_S$, so that, for every $x \in R_S$ and every index $i = 1, \cdots, q$, we have $[z_i, x] = \sum_{j=1}^{p} s_{ij}(x)z_j$, with $s_{ij}(x) \in R_S$. Since the z_j's belong to a free system of generators for R_K over R_S, the coefficients $s_{ij}(x)$ are unique, whence the s_{ij} are linear maps of R_S into itself. Put $Q_{ij} = s_{ij}^{-1}(C)$. Then C' is evidently the intersection of C and all these Q_{ij}. Since R_S/Q_{ij} is isomorphic (via s_{ij}) with a subspace of R_S/C, it is finite dimensional. It follows now by an elementary argument that R_S/C' is also finite dimensional, which was our assertion.

We are now in a position to prove the result alluded to above.

LEMMA 4. *Let K be a finite dimensional Lie algebra over the field F, and suppose that $K = S + L$, where S is a subalgebra, L an ideal, and $S \cap L = (0)$. Let N be a finite dimensional L-module, and suppose that*

$[S, L]$ is contained in an ideal of L which is nilpotent on N. Let M be a K-module which, as an L-module, is contained in N. Then there exists a finite dimensional K-module Z which contains M as a K-submodule and N as an L-submodule.

Proof. Make a composition series $(0) = N_0 \subset N_1 \subset \cdots \subset N_r = N$, such that each factor module N_i/N_{i-1} is L-simple. Let A be the set of all elements of R_L which annihilate each N_i/N_{i-1}. Since $[S, L]$ is contained in an ideal of L which is nilpotent on N, we have $[S, L] \subset A$. Since A contains the annihilator of N in R_L, it is of finite codimension in R_L. As we have seen above, this implies that A^r is also of finite codimension in R_L, and evidently A^r annihilates N. Now let B_1 denote the annihilator of M in $SR_S = R_S{}^+$, and define a decreasing sequence of ideals in R_S by putting (in the notation introduced above) $B_{i+1} = B_i'$, for each $i \geqq 1$. B_1, and hence also each B_i is of finite codimension in R_S. Now let $C = B_r R_L + B_{r-1} A + \cdots + B_1 A^{r-1} + R_S A^r$. One verifies in a straightforward fashion that C is an ideal of R_K. Furthermore, we have a finite set $x_1, \cdots, x_p$ of elements of R_S which span R_S mod. B_r, and a finite set $y_1, \cdots, y_q$ of elements of R_L which span R_L mod. A^r. The products $x_i y_j$ span $R_K = R_S R_L$ mod. $B_r R_L + R_S A^r$ and a fortiori mod. C. Hence C is of finite codimension in R_K. Using that R_S is the direct sum of the subspaces F and $R_S{}^+$, we see from the definition of C that $C \cap R_L = A^r$, and therefore annihilates N.

Now put $T = (R_K/C) \otimes N$, where the tensor product is taken with respect to R_L. T is an R_K-module in the natural fashion, and, since C is of finite codimension in R_K, T is finite dimensional. Let T_0 be the tensor product $(R_K/C) \otimes M$ with respect to R_L. Since C annihilates M, we may regard M as an R_K/C-module, so that there is a canonical homomorphism of T_0 onto M. Let H_0 be the kernel of this homomorphism, and let H be the canonical image of H_0 in T. Clearly, H is an R_K-submodule of T, and we define Z to be the R_K-module T/H. We claim that the map $n \to 1 \otimes n + H$ is an R_L-isomorphism of N into Z, and that its restriction to M is an R_K-isomorphism of M into Z.

In order to prove this, we observe that we have $(C + R_L) \cap (C + R_S{}^+ R_L) = C$. Hence, as a right module over R_L, R_K/C is the direct sum of the submodules $(C + R_L)/C$ and $(C + R_S{}^+ R_L)/C$. By a standard result on tensor products, $(R_K/C) \otimes N$ is therefore canonically isomorphic with the direct sum of the partial tensor products corresponding to our decomposition of R_K/C. Since $C \cap R_L$ annihilates N, the tensor product of $(C + R_L)/C$ by N may be identified with N, and our map of N into Z then reads simply $n \to n + H$. The elements of H are precisely the elements of the form

$m + \sum_i u_i' \otimes m_i$, where m and the m_i belong to M, and the u_i' are the cosets mod. C of elements u_i in $R_S{}^+R_L$ such that $m + \sum_i u_i \cdot m_i = 0$, in M. From this it is clear that $H \cap N = (0)$. Hence we conclude that our map of N into Z is an R_L-isomorphism. Finally, if $m \in M$ and $r \in R_K$, write $r = a + b$, with $a \in R_L$ and $b \in R_S{}^+R_L$. Then we have, $r \cdot (m + H) = r' \otimes m + H = a \cdot m + b' \otimes m + H = a \cdot m + b \cdot m + H = r \cdot m + H$, whence the restriction of our map to M is an R_K-isomorphism. This completes the proof of Lemma 4.

3. Effacing of cohomology classes. A cohomology class c in $H(L, V)$ is called *effaceable* if and only if there exists a finite dimensional L-module U containing V such that the canonical homomorphism $H(L, V) \to H(L, U)$ annihilates c. If c is effaceable, and if U is such a module, we shall say that c is effaced in U, or that U is an effacing module for c. Everything we have done up to now was aimed at the following result.

THEOREM 1. *Let K be a finite dimensional Lie algebra over a field F of characteristic 0, and let M be a finite dimensional K-module. Let S be a maximal semisimple subalgebra of K. Then a cohomology class for K in M is effaceable if and only if it is annihilated by the restriction homomorphism $H(K, M) \to H(S, M)$.*

Proof. Since every finite dimensional S-module is semisimple, the canonical homomorphism $H(S, M) \to H(S, N)$ is an isomorphism whenever N is finite dimensional and $M \subset N$. It follows that the condition of Theorem 1 is necessary.

Now let L denote the radical of K. Then, as is well known, we have $K = S + L$, and $S \cap L = (0)$. By Theorem 13 of [5], there is an isomorphism ϕ_M of the tensor product (with respect to F) $H(S, F) \otimes H(L, M)^K$ onto $H(K, M)$, where $H(L, M)^K$ is the subspace of $H(L, M)$ which consists of all the cohomology classes that are annihilated by K in the canonical structure of $H(L, M)$ as a K-module. Under this isomorphism ϕ_M, the subspace $\sum_{p+q=n} H^p(S, F) \otimes H^q(L, M)^K$ corresponds to $H^n(K, M)$. The isomorphism ϕ_M can be described explicitly in terms of cochains, and it is then easily seen (in [5]) that it has the following properties. If ρ denotes the restriction homomorphism $H(K, M) \to H(S, M)$, then $\rho\phi_M$ is an isomorphism on $H^n(S, F) \otimes H^0(L, M)^K$. Furthermore, if N is a finite dimensional K-module containing M, and if γ denotes the canonical homomorphisms $H(K, M) \to H(K, N)$ and $H(L, M) \to H(L, N)$, we have $\gamma\phi_M(u \otimes v) = \phi_N(u \otimes \gamma(v))$, for all $u \in H(S, F)$ and all $v \in H(L, M)^K$.

Now suppose that $c \in H^n(K, M)$ and $\rho(c) = 0$. Write $c = \phi_M(\sum_{p+q=n} w_{p,q})$, with $w_{p,q} \in H^p(S, F) \otimes H^q(L, M)^K$. If $q > 0$ we have $\rho\phi_M(w_{p,q}) = 0$, by the definition of ϕ_M. Hence $\rho(c) = 0$ implies that $\rho\phi_M(w_{n,0}) = 0$, whence we conclude that $w_{n,0} = 0$. Thus $c = \phi_M(w_{n-1,1} + \cdots + w_{0,n})$. Combining Lemmas 3 and 4, we see immediately that there exists a finite dimensional K-module N containing M such that γ annihilates every $H^q(L, M)$ with $q > 0$. For this module N, we have $\gamma(c) = 0$, by what we have remarked above. This completes the proof of Theorem 1.

It was shown by Iwasawa in [6] that every two dimensional cohomology class for a finite dimensional Lie algebra in a finite dimensional module is effaceable. In characteristic 0, this follows from Theorem 1 and the well known fact that $H^2(S, M) = (0)$. In characteristic p, Iwasawa's result is much easier to prove. For the sake of completeness, we shall give a simplified version of Iwasawa's proof, using the technique of Jacobson, [7].

Let K be a finite dimensional Lie algebra over the field F of characteristic p, let M be a finite dimensional K-module, and f a 2-cocycle for K in M. Regard M as an abelian Lie algebra, and consider the extension (E, ϕ) of M by K which is determined by f. The space of E is the direct sum (K, M) of K and M, the commutation is given by $[(x_1, m_1), (x_2, m_2)] = ([x_1, x_2], x_1 \cdot m_2 - x_2 \cdot m_1 + f(x_1, x_2))$, and $\phi(x, m) = x$. For $e \in E$, denote by D_e the inner derivation of E which is effected by e, so that $D_e(e') = [e, e']$. Then each p^k-th power of D_e is a derivation of E. Let $(x_1, \cdots, x_n)$ be a basis for K and write $D_i = D_{(x_i,0)}$. Then, for each i, some non trivial F-linear combination $\sum_k a_{ik} D_i^{p^k}$ must vanish. Put $u_i = \sum_k a_{ik}(x_i, 0)^{p^k}$. Then each u_i belongs to the center of R_E. Let U be the ideal of R_E which is generated by $u_1, \cdots, u_n$. Put $N = R_{E^+}/(U + MR_{E^+})$. It is easily seen that N is finite dimensional. The left multiplications of R_E induce on N the structure of an R_K-module. We claim that the map $m \to m + U + MR_{E^+}$ is an R_K-isomorphism of M into N. Indeed, for $x \in K$, we have $x \cdot (m + U + MR_{E^+}) = (x, 0)m + U + MR_{E^+} = x \cdot m + U + MR_{E^+}$, whence our map is an R_K-homomorphism. On the other hand, we can write the elements of R_E uniquely as linear combinations of ordered monomials in the elements of a basis of M and the elements $(x_i, 0)$. Each ordered monomial occurring in an element of U must have positive degree in at least one of the $(x_i, 0)$, and each ordered monomial occurring in an element of MR_{E^+} must have total degree at least 2. Hence $M \cap (U + MR_{E^+}) = (0)$, and therefore our map $M \to N$ is an isomorphism.

Now put $g(x) = (x, 0) + U + MR_{E^+}$, for $x \in K$. Then g is a 1-cochain

for K in N, and one checks directly that $\delta g = f$. Thus N effaces the cohomology class of f, and Iwasawa's result is proved also in characteristic p.

Now we recall (cf. [3]) that the cohomology groups of a Lie algebra K can be computed also from the complex of cochains for R_{K^+} in the module M. These cochains are the n-linear functions $(n = 0, 1, \cdots)$ from R_{K^+} to M, and the coboundary operator δ is defined by the formula

$$\delta g(r_0, \cdots, r_n) = r_0 \cdot g(r_1, \cdots, r_n) + \sum_{i=1}^{n} (-1)^i g(r_0, \cdots, r_{i-1} r_i, \cdots, r_n).$$

If Q is an ideal of R_{K^+} which annihilates M, so that M may be regarded also as an R_{K^+}/Q-module, there is an evident canonical homomorphism of $H(R_{K^+}/Q, M)$ into $H(R_{K^+}, M)$. We shall say that a cohomology class c is of *finite type* if it belongs to the canonical image of $H(R_{K^+}/Q, M)$ in $H(R_{K^+}, M) = H(K, M)$, for some ideal Q which is of finite codimension in R_{K^+} and annihilates M. We shall then say also that Q reduces c. We assume throughout in what follows that K and M are finite dimensional. Then it is almost evident that every 0-dimensional cohomology class for K in M is of finite type. Indeed, such a cohomology class is represented by an element of M which is annihilated by R_{K^+}, and we may take for Q simply the annihilator of M in R_{K^+}. Similarly, we see at once that every 1-dimensional cohomology class for K in M is of finite type. In fact, we may take for Q the square of the annihilator of M in R_{K^+}.

Next we observe that, in positive dimensions, every cohomology class which is of finite type is also effaceable. In fact, if Q is an ideal of R_{K^+} which reduces the cohomology class c of positive dimension, the module $\mathrm{Hom}_F(R_K/Q, M)$ of all F-linear maps of R_K/Q into M is an effacing module for c.

THEOREM 2. *Every cohomology class of dimension* 0, 1, *or* 2 *is of finite type, and every effaceable cohomology class of dimension* 3 *is of finite type.*

Proof. We have already disposed of the trivial cases of dimensions 0 and 1. Now let c be a two dimensional cohomology class for K in M. We have seen above that there is a finite dimensional effacing module N for c. Let Q be the annihilator of N in R_{K^+}, and let f be a 2-cocyle for R_{K^+} in M which represents c. There is a 1-cochain g for R_{K^+} in N such that $f(r, s) = r \cdot g(s) - g(rs)$, for all r and s in R_{K^+}. Taking r in Q, we see that $g(QR_{K^+}) \subset M$. Take a 1-cochain h for R_{K^+} in M which coincides with g on QR_{K^+}. Then the cocycle $f_1 = f - \delta h$ still represents c. We have $f_1(Q, R_{K^+}) = (0)$, and a fortiori $f_1(Q^2, R_{K^+}) = (0)$. Using also that f_1 is a

cocycle, we find that $f_1(R_K^+, Q^2) = (0)$. Hence c belongs to the canonical image of $H(R_K^+/Q^2, M)$. Since Q is of finite codimension, so is Q^2, and we have shown that c is of finite type.

Finally, suppose that c is of dimension 3 and is effaced in the finite dimensional K-module N containing M. Then c is the image of a cohomology class e in $H^2(K, N/M)$, under the coboundary map of the exact sequence for the pair (N, M). As we have just shown, e is of finite type. Let P be an ideal of finite codimension in R_K^+ which reduces e. Then we can find a 2-cochain g for R_K^+ in N which vanishes on $P \otimes R_K^+ + R_K^+ \otimes P$ and which, when taken mod. M, represents e. Its coboundary δg takes values in M and represents c. Let Q be the annihilator of N in R_K^+. Then it is evident that $\delta g(u, v, w) = 0$ whenever one of u, v, or w lies in $P \cap Q$. Hence $P \cap Q$ reduces c, and Theorem 2 is proved.

Our last argument can easily be generalized to prove the following result.

THEOREM 3. *Let K be a finite dimensional solvable Lie algebra over a field of characteristic 0, and let M be a finite dimensional K-module. Then every cohomology class for K in M is of finite type.*

Proof. We merely have to observe that every cohomology class for K in M which is of positive dimension is effaceable, and then prove our result by induction on the dimension, the inductive step being exactly like our above proof for the case of dimension 3.

Finally, we shall show by means of an example that *not every effaceable cohomology class is of finite type.* Let S be a semisimple Lie algebra over a field F of characteristic 0, and let Z be the one dimensional Lie algebra over F. Put $L = S \oplus Z$ (direct sum). We shall exhibit a 4-dimensional effaceable cohomology class for L in F which is not of finite type. It is known that $H^3(S, F) \neq (0)$. Let f be a 3-cocycle for R_S^+ in F which is not a coboundary. We can evidently extend f to a 3-cocycle for R_L^+ in F, still denoted f, such that $f(a, b, c) = 0$ whenever one of the arguments a, b, or c lies in ZR_L. On the other hand, let g be the 1-cocycle for R_L^+ in F such that, for some fixed non-zero element z of Z, $g(z) = 1$, while $g(R_S^+ + ZR_L^+) = (0)$. Now define $h(a, b, c, d) = g(a)f(b, c, d)$. Then h is clearly a 4-cocycle for R_L^+ in F. Since the restriction of h to R_S^+ is 0, it follows at once from Theorem 1 that the cohomology class of h is effaceable. We shall show that the cohomology class of h is not of finite type. In fact, suppose that it is of finite type, and let P be an ideal of finite codimension in R_L^+ which reduces it. Then there exists a 3-cochain k for R_L^+ in F such that $(h - \delta k)(a, b, c, d) = 0$ whenever one of a, b, c, or d belongs to P.

Taking $a = z$, we find that we have

$$f(b, c, d) + k(zb, c, d) - k(z, bc, d) + k(z, b, cd) = 0$$

whenever one of b, c, or d belongs to P. On the other hand, using that $h(a, b, c, d) = 0$ whenever one of b, c, or d is equal to z, and that z is in the center of R_L, we see that, whenever one of b, c, or d belongs to P, we have

$$\begin{aligned} k(zb, c, d) &= k(bz, c, d) = k(b, zc, d) - k(b, z, cd) = k(b, cz, d) - k(b, z, cd) \\ &= k(bc, z, d) + k(b, c, zd) - k(b, z, cd) = k(bc, z, d) + k(b, c, dz) - k(b, z, cd) \\ &= k(bc, z, d) + k(b, cd, z) - k(bc, d, z) - k(b, z, cd). \end{aligned}$$

Substituting this above, we find that $f(b, c, d) - m(bc, d) + m(b, cd) = 0$ whenever one of b, c, or d belongs to P, where

$$m(u, v) = k(z, u, v) - k(u, z, v) + k(u, v, z).$$

Now the restriction to $R_S{}^+$ of $f + \delta m$ represents the same cohomology class in $H^3(S, F)$ as f. The last result shows that this cohomology class is reduced by $P \cap R_S{}^+$ which is of finite codimension in $R_S{}^+$. Hence this cohomology class is effaceable, which is a contradiction because it is not 0, and S is semisimple.

4. Finite dimensionality of kernels. Let M be a Lie algebra over the field F. Denote by $D(M)$ the Lie algebra consisting of the derivations of M, and by $I(M)$ the ideal of $D(M)$ consisting of the inner derivations. If K is another Lie algebra over F we say that we have the structure of a K-kernel on M if we are given a homomorphism ψ of K into $D(M)/I(M)$. A linear map $x \to t_x$ of K into $D(M)$ which becomes ψ when reduced mod. $I(M)$ is called an *operator map* for the kernel (M, ψ). We can find a bilinear alternating map τ from K to M such that $[t_x, t_y] - t_{[x,y]}$ is the inner derivation of M which is effected by $\tau(x, y)$, for all x and y in K. Such a map τ is called a *hindrance* belonging to the kernel (M, ψ). The formal coboundary

$$\begin{aligned} \delta\tau(x, y, z) = {} & t_x(\tau(y, z)) - t_y(\tau(x, z)) + t_z(\tau(x, y)) \\ & - \tau([x, y], z) + \tau([x, z], y) - \tau([y, z], x) \end{aligned}$$

lies in the center N of M, on which there is an induced structure of a K-module (the operators being given by the t_x). Furthermore, $\delta\tau$ is actually a 3-cocycle for K in N, and its cohomology class in $H^3(K, N)$ depends only on the kernel (M, ψ), and not on our particular choice of t and τ. This cohomology class is called the *obstruction* of our kernel (M, ψ). The kernel

(M, ψ) is said to be *extendible* if it arises in the natural fashion from a Lie algebra extension of M by K, which is the case if and only if its obstruction is 0. One can define an operation of addition of kernels with the same center N and then partition these kernels into equivalence classes, two kernels M_1 and M_2 being equivalent if and only if there are extendible kernels U_1 and U_2 such that $M_1 + U_1 = M_2 + U_2$. The equivalence classes of kernels with a fixed center N can be shown to constitute a vector space over F (the zero element being the class of the extendible kernels) such that our above association of an element of $H^3(K, N)$ with a kernel (M, ψ) induces an isomorphism of the space of the kernel classes into $H^3(K, N)$. Moreover, we have shown in [4] that this isomorphism is *onto* by giving a construction which attaches to a given 3-cocycle for R_{K^+} in N a kernel (M, ψ) with center N whose obstruction coincides with the cohomology class of the given 3-cocycle. If the given 3-cocycle is the canonical image of a 3-cocycle for R_{K^+}/Q in N, where Q is an ideal of finite codimension in R_{K^+} and annihilating N, we can carry out our standard construction with the algebra R_K/Q in the place of R_K. This yields (assuming that K and N are finite dimensional) a kernel (M, ψ) which is finite dimensional and still has the same obstruction. Hence we have the following result.

THEOREM 4. *Let K be a finite dimensional Lie algebra over a field F, and let N be a finite dimensional K-module. Then every cohomology class in $H^3(K, N)$ which is of finite type, or, equivalently, which is effaceable, is the obstruction of a finite dimensional K-kernel with center N.*

It seems likely that the converse of this result is also true. However, we have succeeded in proving this only in the case of characteristic 0. In this case, we have actually the following result.

THEOREM 5. *Let K be a finite dimensional Lie algebra over the field F of characteristic 0, and let N be a finite dimensional K-module. Let S be a maximal semisimple subalgebra of K. Then an element $H^3(K, N)$ is the obstruction of a finite dimensional K-kernel if and only if its restriction to S is 0.*

Proof. If $c \in H^3(K, N)$ and if its restriction to S is 0 then c is effaceable, by Theorem 1. Hence, by Theorem 2, c is of finite type. By Theorem 4, c is therefore the obstruction of a finite dimensional K-kernel.

Conversely, suppose that c is the obstruction of a finite dimensional K-kernel (M, ψ). Then $\psi(S)$ is a semisimple subalgebra of $D(M)/I(M)$. There is a subalgebra E of $D(M)$ which contains $I(M)$ and for which

$E/I(M) = \psi(S)$. Since every finite dimensional extension by a semisimple Lie algebra is trivial (which is essentially a consequence of the vanishing of the 2-dimensional cohomology groups of semisimple Lie algebras), we can find an isomorphism ρ of $\psi(S)$ into E (and thus into $D(M)$) which is inverse to the canonical homomorphism of E onto $\psi(S)$. Now we can evidently find an operator map $x \to t_x$ which coincides with $\rho\psi$ on S. To this, we can find a hindrance τ which vanishes on $S \otimes S$. Then $\delta\tau$ vanishes on $S \otimes S \otimes S$, whence it is clear that the restriction of c to S is 0.

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] S. Eilenberg and S. MacLane, "Cohomology theory in abstract groups, II," *Annals of Mathematics*, vol. 48 (1947), pp. 326-341.

[2] G. Hochschild, "Lie algebras and differentiations in rings of power series," *American Journal of Mathematics*, vol. 72 (1950), pp. 58-80.

[3] ———, "Cohomology of restricted Lie algebras," *ibid.*, vol. 76 (1954), pp. 555-580.

[4] ———, "Lie algebra kernels and cohomology," *ibid.*, vol. 76 (1954), pp. 698-716.

[5] ——— and J-P. Serre, "Cohomology of Lie algebras," *Annals of Mathematics*, vol. 57 (1953), pp. 591-603.

[6] K. Iwasawa, "On the representation of Lie algebras," *Japanese Journal of Mathematics*, vol. 19 (1948), pp. 405-426.

[7] N. Jacobson, "A note on Lie algebras of characteristic p," *American Journal of Mathematics*, vol. 74 (1952), pp. 357-359.

[8] J. L. Koszul, "Sur les modules de réprésentations des algèbres de Lie résolubles," *ibid.*, vol. 76 (1954), pp. 535-554.

[9] G. F. Leger, "On cohomology theory for Lie algebras," Thesis, University of Illinois, 1951.

[10] H. Zassenhaus, "Ueber die Darstellungen der Lie-Algebren bei Charakteristik 0," *Commentarii Mathematici Helvetici*, vol. 26 (1952), pp. 252-274.

SUMS OF DISTINCT DIVISORS.*

By B. M. STEWART.

1. Introduction. For a given positive integer M, let $\alpha(M)$ indicate the number of integers n which can be written in the form

$$n = \sum d \tag{1}$$

where the d are distinct positive divisors of M. Obviously $\alpha(M) \leqq \sigma(M)$, the sum of all the divisors of M.

For a given set S of integers M, define S^* to be the subset of S containing those integers for which $\sigma(M) - \alpha(M)$ is minimal.

Thus for the set I of all positive integers, the set I^* contains all integers, such as $M = 2^k$, for which $\sigma(M) - \alpha(M) = 0$. However, for the set O of all odd integers $M \geqq 5$, the set O^* contains all integers, such as $M = 945$, for which $\sigma(M) - \alpha(M) = 2$, for certainly both 2 and $\sigma(M) - 2$ defy representation in the form (1).

In this paper we give complete structure theorems for I^* and O^* and indicate their application to problems of the Egyptian fraction type. We also show that the function $\alpha(M)/\sigma(M)$ is everywhere dense on the interval 0 to 1.

2. Structure theorems for I^*. It is clear that $M = 2^k$ belongs to I^*, because any n in the range $1 \leqq n \leqq \sigma(M) = 2^{k+1} - 1$ when written in the binary notation is in the form (1).

THEOREM 1. *If M belongs to I^* and p is a prime with $(p, M) = 1$, then a necessary and sufficient condition that $M' = p^k M$, $k \geqq 1$, belong to I^* is that $p \leqq \sigma(M) + 1$.*

Proof. The smallest divisor of M' not a divisor of M is p. If $p > \sigma(M) + 1$, then $n = \sigma(M) + 1$ defies representation with respect to M'.

Conversely, suppose $p \leqq \sigma(M) + 1$. We show by induction that $p^k \leqq \sigma(p^{k-1}M) + 1$. This is obvious for $k = 1$. Using the induction hypothesis on k, we have

* Received May 23, 1953, revised January 28, 1954. Presented to the American Matematical Society, April 25, 1953.

$$p^{k+1} \leqq p\sigma(p^{k-1}M) + p \leqq p\sigma(p^{k-1}M) + \sigma(M) + 1 = \sigma(p^kM) + 1.$$

We assume as an induction hypothesis that $p^{k-1}M$ belongs to I^*. To show that $M' = p^kM$ belongs to I^*, we consider the intervals ranging from Qp^k to $Qp^k + \sigma(p^{k-1}M)$ for $Q = 0, 1, \cdots, \sigma(M)$. It follows that no integer n in the range $1 \leqq n \leqq \sigma(M')$ is omitted from all these intervals. For on the one hand, because $p^k \leqq \sigma(p^{k-1}M) + 1$, we have

$$(Q+1)p^k \leqq Qp^k + \sigma(p^{k-1}M) + 1,$$

hence the intervals are overlapping or contiguous. On the other hand, the intervals include 1 and $p^k\sigma(M) + \sigma(p^{k-1}M) = \sigma(M')$. Thus each n can be written

$$n = Qp^k + R, \qquad 0 \leqq Q \leqq \sigma(M), \qquad 0 \leqq R \leqq \sigma(p^{k-1}M).$$

Since M and $p^{k-1}M$ are assumed in I^*, we can write $Q = \sum d$, where the d are distinct divisors of M, and we can write $R = \sum D$, where the D are distinct divisors of $p^{k-1}M$. Then

$$n = \sum d' + \sum D$$

where the $d' = p^kd$ are distinct from the D because of involving the factor p^k and both the d' and the D are divisors of M'. Hence M' belongs to I^*.

COROLLARY 1. *M belongs to I^* if and only if M is of the form $M = 2^a$, $a \geqq 0$, or of the form*

$$M = 2^a p_1^{a_1} \cdots p_k^{a_k}, \quad a \geqq 1, \quad a_j \geqq 1, \quad j = 1, 2, \cdots, k,$$

where the primes p_j satisfy the following conditions:

$$2 < p_1 < \cdots < p_k, \qquad p_1 \leqq \sigma(2^a) + 1, \text{ and}$$

$$p_{j+1} \leqq \sigma(2^a p_1^{a_1} \cdots p_j^{a_j}) + 1, \qquad j = 1, 2, \cdots, k-1.$$

Proof. It is easy to see that 1 is the only odd number belonging to I^*. We have noted that $M = 2^a$ belongs to I^*. The corollary then follows from Theorem 1 by induction on k.

3. Practical numbers. A. K. Srinivasan [1] has defined a *practical* number M (i. e., a number of advantage for use in weights, measures, or coinage), requiring that M have property (1) for all n with $1 \leqq n \leqq M$. But he failed to find a complete structure theorem for these numbers.

It is obvious that every number belonging to I^* is a practical number. We show below that the converse is true. It follows that Corollary 1 answers

Srinivasan's question about the complete structure theorem for practical numbers.

To establish that every practical number belongs to I^*, we simply review the proof of Theorem 1, noting that if M is assumed practical, the criterion of Theorem 1 applies to decide whether $M' = p^k M$ is practical. For if $p > \sigma(M) + 1$, then not only does $n = \sigma(M) + 1$ defy representation in the form (1) relative to M', but also $n < M'$. Thus the condition of Theorem 1 is necessary for M' to be practical. Since the condition is sufficient to make M' belong to I^*, it is sufficient to make M' practical.

4. Values of $\alpha(M)/\sigma(M)$. We need the following lemma.

LEMMA 1. *If p is a prime such that $p > \sigma(T)$, then*

$$\alpha(pT) = \alpha(T)\{\alpha(T) + 2\}.$$

Proof. Every sum of divisors of pT must have the form $n = n_1 p + n_2$ where n_1 and n_2 are either 0 or a sum of divisors of T. Furthermore, since the maximum value of n_2 is $\sigma(T)$ and $p > \sigma(T)$, no two of these n can be equal unless they have the same n_1 and n_2. Since there are $\alpha(T) + 1$ choices for each of n_1 and n_2, except that the choice 0, 0 must be omitted, we have

$$\alpha(pT) = \{\alpha(T) + 1\}^2 - 1 = \alpha(T)\{\alpha(T) + 2\}.$$

We define $s(M) = \alpha(M)/\sigma(M)$.

THEOREM 2. *The values of $s(M)$ are everywhere dense on the interval* 0 *to* 1.

Proof. For given real numbers x and y such that $0 < x < y \leqq 1$, we must produce an integer K such that $x < s(K) < y$. We try to find $K = pM$ where M is in I^* and $p > \sigma(M)$, for then by Lemma 1

$$s(K) = s(pM) = \sigma(M)\{\sigma(M) + 2\}/(p + 1)\sigma(M) = \{\sigma(M) + 2\}/(p + 1).$$

If we set $u = 1/y$, $u(1 + \epsilon) = 1/x$, our problem is to find M and p so that

$$u\{\sigma(M) + 2\} < p + 1 < u\{\sigma(M) + 2\}(1 + \epsilon).$$

On the one hand we know from Theorem 1 that there exist arbitrarily large numbers M belonging to I^*. On the other hand we have the Cahen-Stieltjes theorem that for any $\epsilon > 0$ and for sufficiently large $v = u\{\sigma(M) + 2\}$ there must exist at least one prime p in the interval $v - 1 < p < (v - 1)(1 + \epsilon)$, hence such that $v < p + 1 < v(1 + \epsilon)$. Since $u \geqq 1$, such a prime p has the required property $p > \sigma(M)$, which completes the proof.

It seems difficult to decide whether $s(M)$ will take on a specified rational value. Even $s(M) = 1/3$ requires some search before we find a solution such as $M = 12 \cdot 89$. For $s(M) = 4/5$ the simplest solution we know is $M = 2^8 \cdot 73 \cdot 47269$.

5. Structure theorem for O^*. A necessary condition that M belongs to O^* is that M be a multiple of 15, for unless M has the factors 3 and 5, the numbers $n = 3$ and $n = 5$ will be exceptions to (1).

Arrange the τ divisors d_j of M in natural order

$$d_1 = 1 < d_2 = 3 < d_3 = 5 < \cdots < d_\tau = M,$$

and let σ_i denote the sum of the first i of the d_j.

THEOREM 3. *Necessary and sufficient conditions that $M = 15T$ belong to 0^* are that for $i \geqq 3$,*

either (a) $\quad d_{i+1} \leqq \sigma_i - 2, \quad d_{i+1} \neq \sigma_i - 4;$

or (b) $\quad d_{i+1} = \sigma_i - 4, \quad d_{i+2} = \sigma_i - 2.$

Proof. For numbers of O, it is clear that 2 does not have the form (1). Because of the ordering of the d_j, it follows that if $d_{i+1} > \sigma_i - 2$, then $n = \sigma_i - 2$ does not have the form (1). If $d_{i+1} = \sigma_i - 4$, then $n = \sigma_i - 2 = d_{i+1} + 2$ again defies representation unless $d_{i+2} = \sigma_i - 2$. Therefore the conditions (a) or (b) are necessary when $i \geqq 3$ if M is to belong to O^*.

Conversely, suppose (a) or (b) satisfied when $i \geqq 3$. As an induction hypothesis on i assume when $3 \leqq i \leqq \tau - 1$ and $2 < x < \sigma_i - 2$ that either $x = \sum d$ using distinct $d \leqq d_i$ or that $x = \sigma_{i-1} - 2 = d_i + 2 = d_{i+1}$. This hypothesis is easily verified when $i = 3$.

Consider y where $2 < y < \sigma_{i+1} - 2$. When $y < \sigma_i - 2$ set $y = x$ and use the induction hypothesis to write $y = \sum d$ where $d \leqq d_{i+1}$. When $\sigma_i - 2 \leqq y < \sigma_{i+1} - 2$ set $y = d_{i+1} + x$ and note from (a) or (b) that $0 \leqq x < \sigma_i - 2$. Except when $x = 2$ or $x = d_{i+1} = d_i + 2$, it follows from the induction hypothesis that $y = \sum d$ with distinct $d \leqq d_{i+1}$.

When $x = d_{i+1} = d_i + 2 = \sigma_{i-1} - 2$ we find that $y = d_{i+1} + x = d_i + \sigma_{i-1} = \sigma_i$. When $x = 2$ the conditions $\sigma_i - 2 \leqq y = d_{i+1} + 2$ and $d_{i+1} \leqq \sigma_i - 2$ imply that either $y = \sigma_i$, $\sigma_i - 1$, or $\sigma_i - 2$. The first two cases obviously allow $y = \sum d$ with $d \leqq d_{i+1}$. The last case $y = d_{i+1} + 2 = \sigma_i - 2$ because of (b) has the representation $y = d_{i+2}$.

Thus we have established for the case $i + 1$ the exact counterparts of the induction hypothesis on i. It follows that every n satisfying

$2 < n < \sigma(M) - 2$ has a representation (1); so do $n = 1$, $\sigma(M) - 1$, and $\sigma(M)$, hence M belongs to O^*.

THEOREM 4. *If M belongs to O^* and p is an odd prime, then $M' = pM$ belongs to O^* if and only if*

$$2p \leqq \sigma(M) - 2, \qquad 2p \neq \sigma(M) - 4.$$

Proof. Let the divisors and partial sums of M' be indicated by d_i' and σ_i', respectively, and note that the only divisors of M' which are not divisors of M are multiples of p^{a+1} where $M = p^a A$, $a \geqq 0$, $(p, A) = 1$.

Supose that the first t divisors of M' coincide with those of M, so that $d_i = d_i'$, $i \leqq t$, $3 \leqq t$. Then the properties (a) or (b) of M as described in Theorem 3 carry over to M' for $3 \leqq i < t$.

Next consider cases $t \leqq r < k$ where $d_r \leqq d_k' < d_{k+1}' \leqq d_{r+1}$. By Theorem 3 applied to M we have

$$d_{k+1}' \leqq d_{r+1} \leqq \sigma_r - 2 < \sigma_r + p^{a+1} - 4 \leqq \sigma_k' - 4,$$

hence case (a) holds for M' and k.

Finally, consider cases where $M \leqq d_k' < d_{k+1}' \leqq M'$. Then $d_{k+1}' = M'/d_j$ where $d_j < p$ hence d_j divides M as the notation indicates. When $M/d_j > 5$, it follows from Theorem 3 applied to M for $i \geqq 3$ that

$$M/d_j \leqq \sigma(M) - M/d_1 - M/d_2 - \cdots - M/d_j - 2.$$

Hence

$$\begin{aligned} d_{k+1}' = pM/d_j &\leqq p\sigma(M) - M'/d_1 - M'/d_2 - \cdots - M'/d_j - 2p \\ &< \sigma(A) + p\sigma(M) - M'/d_1 - M'/d_2 - \cdots - M'/d_j - 4 = \sigma_k' - 4, \end{aligned}$$

so that case (a) holds for M' and k. There remain only the possibilities that $d_{k+1}' = p$, $3p$, or $5p$ with $M < d_{k+1}'$. In case $3p < M < 5p$ we have $p < M/3 < \sigma(M) - 4$, hence $d_{k+1}' = 5p < \sigma(M) + p + 3p - 4 = \sigma_k' - 4$ so that (a) holds for M' and k. In case $p < M < 3p$ or $M < p$, the condition $d_{k+1}' = 3p \leqq \sigma_k' - 2 = \sigma(M) + p - 2$ is necessary for M' to belong to O^*. The contingency suggested by (b) is ruled out for it would require $d_{k+1}' = 3p = d_{k+2}' - 2 = 5p - 2$, hence case (a) must hold with $3p \neq \sigma(M) + p - 4$. Thus the conditions of Theorem 4 are necessary. Conversely, when these conditions are satisfied, then

$$d_{k+1}' = 5p < \sigma(M) + p + 3p - 4 = \sigma_k' - 4;$$

and if $M < p$, the condition $d_{k+1}' = p < \sigma(M) - 4 = \sigma_k' - 4$ holds.

Therefore for all $k \geqq 3$ the conditions of Theorem 3 hold and M' belongs to O^*.

COROLLARY 4. *If the prime p divides M and M belongs to O^*, then $M' = pM$ belongs to O^*.*

Proof. We have noted that $M = 15T$ hence when $p > 5$, we have $\sigma(M) > 15p > 2p + 4$; and if $p = 3$, or 5, we have $\sigma(M) > 15 > 2p + 4$; so the conditions of Theorem 4 hold.

We cannot give any simple catalog for O^* like that for I^* in Corollary 1, because there are infinitely many different types of "base" numbers in O^* instead of the single type 2^a noted for I^*. However, a few examples will help fix ideas.

Since $M = 15T$ has $\sigma_3 = 9$, it follows from Theorem 3 that if M is in O^*, then $d_4 = 7$. This implies $\sigma_4 = 16$ and hence that $d_5 = 9$, 11, or 13. It happens that all of $3^2 \cdot 5 \cdot 7$, $3 \cdot 5 \cdot 7 \cdot 11$, and $3 \cdot 5 \cdot 7 \cdot 13$ fail to belong to O^*, but only because $d_\tau > \sigma_{\tau-1} - 2$. It is then easy to check, for example, that $M = 3^2 \cdot 5 \cdot 7p$ is in O^* for any odd prime $p \leqq 103$. In particular, when $p = 3$, we find $M = 945$, the smallest integer in O^*. For use in the sequel, we apply Theorem 4 to see that $M = 3^a \cdot 5 \cdot 7$ is in O^* for $a \geqq 3$.

Another special case of interest is that if M_k is the product of the first k odd primes, then M_k is in O^* when $k \geqq 5$.

The author is indebted to the referee for noting and insisting on possibility (b) in Theorem 3, despite the fact that the smaller numbers in O^* consistently satisfy (a). By considering the arguments of Theorems 3 and 4, the author has constructed the following example: $M = 3^2 \cdot 5^2 \cdot 7^2 \cdot 229pq$ where prime $p = 2{,}641{,}663$ and prime $q = 7{,}924{,}991$, such that M is in O^* and has $d_{56} = 3p = \sigma_{55} - 4$ and $d_{57} = q = \sigma_{55} - 2$, so that case (b) occurs for $i = 55$.

6. Egyptian fraction problems. Let a positive rational number with numerator 1 be called a unit fraction. Then the usual Egyptian fraction problem is to show that any given positive rational number x may be written as a sum of a finite number of distinct unit fractions. A variation of this problem proposed by E. P. Starke [2] is to show that if x has an odd denominator, then x may be written as a sum of a finite number of distinct unit fractions with odd denominators.

Whether $u_n = 1/n$ or $u_n = 1/(2n + 1)$, the sequence $\{S_t\}$, where $S_t = u_1 + u_2 + \cdots + u_t$, is divergent, hence it is easy to replace the problem of representing an improper fraction x by that of representing a proper fraction $A/B = x - S_t$ where $A/B < u_{t+1}$ is so small that unit fractions used in its representation cannot duplicate those used in the partial sum S_t.

For the usual Egyptian fraction problem easy constructions are known, but an amusing variant is provided by our knowledge of numbers belonging to I^*. By Corollary 1 we can find C, one choice being $C = 2^a$ for a sufficiently large a, so that $BC = M$ will belong to I^*. Since $AC < M$, we may write $AC = \sum d$, using distinct divisors d of M. If we set $dd' = M$, then

$$A/B = AC/BC = \sum d/M = \sum 1/d'$$

solves the problem.

For Starke's problem, our knowledge of numbers belonging to O^* provides the same sort of solution. We can find C, one choice being $C = 3^a(945)$ for a sufficiently large a, so that $BC = M$, by Theorem 4, will belong to O^* and so that $2 < AC < M$, and then proceed as explained above.

MICHIGAN STATE COLLEGE.

REFERENCES.

[1] A. K. Srinivasan, "Practical numbers," *Current Science* (1948), pp. 179-180.

[2] E. P. Starke, "Advanced problem 4512," *American Mathematical Monthly*, vol. 59 (1952), p. 640.

PATHS AND CIRCUITS IN CRITICAL GRAPHS.*

By JOHN B. KELLY and L. M. KELLY.

---

1. Introduction. An admissible coloring of a graph is an assignment of colors to the nodes in such a way that no two nodes which are joined are assigned the same color. If, in coloring a graph G, k is the minimum number of colors which may be employed, then G is said to be k-chromatic. In analyzing k-chromatic graphs one is led naturally to a consideration of those graphs which 'just barely' need k colors. Specifically, a graph is called critically k-chromatic or k-critical if any proper subgraph can be colored in less than k colors.

Each k-critical graph of order n has associated with it a maximal circuit length. Let the minimum of all such maximal circuit lengths for a fixed n and k be denoted by $L_k(n)$.[1] A number of theorems and conjectures advanced by Dirac [1] may be conveniently formulated in terms of this function. Thus, for example, he has shown that $L_k(n) \geqq 2k-2$ provided $k \geqq 3$ and $n \geqq 2k-2$. We shall be interested here in the behavior of $L_k(n)$ for large values of n. Dirac has conjectured that:

(i) $\lim_{n\to\infty} L_k(n) = \infty$ and (ii) $\liminf L_k(n)/n^{\frac{1}{2}} > 0, k \geqq 3.$

In section 3 of this paper we prove that (i) is true while in section 4 we show that (ii) is false when $k = 4$ and in fact that

$$\liminf L_4(n)/\log^2(n) \leqq 3/\log^2(27/4).$$

We believe that (ii) is false for all values of k but have been unable to establish this. The determination of the true order of magnitude of $L_k(n)$ even for $k = 4$ seems to be a difficult problem.

Let $S_k(n)$ denote the function dual to $L_k(n)$, that is, the maximum of the minimal circuit lengths contained in the critical k-chromatic graphs with n nodes, n and k fixed. About this function we know even less than we know about $L_k(n)$. We conjecture that $\limsup S_k(n) = \infty$, $k \geqq 3$, but have been able to prove only that l. u. b. $S_k(n) \geqq 6$, $k \geqq 3$ or in other words, that there exist for any integer k, $k \geqq 3$, k-chromatic critical graphs with no circuits of length less than 6. This is demonstrated in section 5.

---

* Received March 2, 1954.

[1] Relatively simple examples show that the function $L_k(n)$ is defined for all n and k, $k > 3$, $n > k-1$, $n \neq k+1$.

2. Preliminaries. Definitions and Notation.

Definition 2.1. A *graph* G is a collection of elements N called variously *nodes, vertices, points,* etc., and a collection E of unordered pairs of these elements. If (v_1, v_2) is an element of E, then we say that v_1 and v_2 are joined and the pair is called an *edge.*

Definition 2.2. A subgraph of G is a subset M of N together with those edges (x, y) of E such that $x \in M$ and $y \in M$. We note that every subset of the nodes N determines a subgraph.

Definition 2.3. A k-coloring of G, k a positive integer, is a mapping f of N into the integers $1, 2, \cdots, k$ such that if $(v_1, v_2) \in E$, then $f(v_1) \neq f(v_2)$. A graph G is called k-chromatic if it admits a k-coloring but no $k-1$ coloring.

Definition 2.4. A graph G is called *critically k-chromatic* or simply *k-critical* if it is k-chromatic but each of its proper subgraphs is r-chromatic, $r < k$. It is easy to show that if a k-critical graph has n nodes, then any of the subgraphs with $n-1$ nodes is $k-1$ chromatic.

Definition 2.5. The *order* of a graph is the number of its nodes. In this paper all graphs are of finite order. The *degree* of a node is the number of edges to which it belongs.

Definition 2.6. A *path* in a graph G is a set of nodes $v_1 v_2, \cdots, v_m$ such that $(v_1 v_2), (v_2, v_3), \cdots, (v_{m-1}, v_m)$ are all edges, together with those edges. The number of nodes is the *length* of the path.

Definition 2.7. In the definition of a path it is understood that the nodes are distinct. If however we allow the first and last nodes to be coincident we then refer to the path as a *circuit.* The length of a circuit is the number of its nodes.

Definition 2.8. A graph is *connected* if each two of its points may be joined by some path. A connected subgraph, not a proper subgraph of any connected subgraph of G, is a *component* of G. That is, a component is a maximal connected subgraph.

Definition 2.9. A set S of nodes of a connected graph G is called a *cut set* of G, if $G - S$ is not connected. Here $G - S$ is the subgraph defined by the nodes of G which are not in S. Throughout the paper we shall use this notation. Thus $A + B$ will be the subgraph of G defined by the set of nodes A and the set of nodes B. The *index* of a cut set S relative to G is the number of components in the graph $G - S$.

Definition 2.10. A graph G is *Hamiltonian* if there is a path in G including all the nodes of G.

3. The proof of conjecture (*i*).

LEMMA 3.1. *If f is a k-coloring of a graph G, and P is a subset of p nodes of G, then there exists a k-coloring F of G such that $1 \leqq F(x) \leqq p$, $x \in P$.*

Proof. If $p \geqq k$ there is nothing to prove. If $p < k$ the lemma follows by a simple interchange of functional values.

THEOREM 3.1. *If P is a cut set of p points and of index r of a k-critical graph G, then $r < p^p$.*

Proof. Suppose $C_1, C_2, \cdots, C_r$ the components of $G - P$. Since $G - C_1$ is colorable in $k-1$ colors, it follows that there is a coloring function F defined on $G - C_1$ with positive integral values less than k. Furthermore this function may be so chosen that $1 \leqq F(x) \leqq p$ for $x \in P$. This function F in turn induces coloring functions $f_{12}, f_{13}, f_{14}, \cdots, f_{1r}$ defined on $C_2 + P, C_3 + P, \cdots, C_r + P$ respectively, such that $1 \leqq f_{1i} \leqq k-1$, $1 \leqq f_{1i}(x) \leqq p$, $x \in P$ and $f_{12}(x) = f_{13}(x) \cdots = f_{1r}(x)$, $x \in P$. Similarly for $G - C_2$, etc. The resulting functions may be conveniently exhibited thus:

$$\begin{array}{c} -f_{12}f_{13}\cdots f_{1r} \\ f_{21}-f_{23}\cdots f_{2r} \\ \cdots\cdots\cdots \\ f_{r1}f_{r2}\cdots - . \end{array}$$

Clearly if r exceeds the maximum number of ways of coloring p points with p colors, namely p^p, then the functional values of the functions in at least two of the rows, say 1 and 2, must coincide on P. That is $f_{12}(x) = f_{13}(x) \cdots = f_{1r}(x) = f_{21}(x) = f_{23}(x) = \cdots f_{2r}(x)$ for all x in P. Thus if we define

$$f(x) = \begin{cases} f_{21}(x), x \in C_1 + P \\ f_{1i}(x), x \in C_i, i \neq 1 \end{cases}$$

then the result is a $k-1$ coloring of G contrary to the hypothesis. Therefore $r < p^p$.

This result can easily be improved to $r < p^{p-1}$. From this we conclude that a k-critical graph has no cut points and that a cut pair in a critical graph is of index 2.

THEOREM 3.2 (Dirac). *If in a connected graph G, the maximal degree of the nodes is v and the length of the longest path is l, then the order of G cannot exceed $v + v^2 + v^3 \cdots v^l$.*

COROLLARY 3.2.1. *If a graph G has c components, the maximal degree of the nodes is v, and the length of the longest path in G is l, then the order of G cannot exceed $c(v + v^2 + \cdots + v^l)$.*

Definition 3.2. The *extension* of a path $P = p_1p_2p_3 \cdots p_r$ is the subgraph defined by the collection of nodes p_x of G such that p_x is an element of a path $p_1p_2p_3 \cdots p_r \cdots p_x$ but p_x is not an element of P. We designate the extension of P by P^*.

LEMMA 3.2. *A component of P^* is a component of $G - P$.*

Proof. This follows at once from the fact that a node of P^* and a node of $G - (P + P^*)$ cannot be joined by a path which contains no points of P. It is of interest to observe that a path P will be a cut set of G if neither $G - (P + P^*)$ nor P^* is empty.

THEOREM 3.3. $\lim_{n \to \infty} L_k(n) = \infty$.

Proof. Let G be a k critical graph with maximal path length l and classify the points of G as follows. A point p is of type i provided it is the i-th point of some path and i is the largest number for which this is true. Clearly every point of G is in some unique class $1, 2, 3, \cdots, l$. First observe that points of type l have degree less than l. This will serve as an anchor for a (decreasing) induction.

Suppose now that nodes of type x, $i \leqq x \leqq l$ have maximal degree $O_i(l, k)$, where O_i is a bounded function of l and k, and consider a point p_{i-1} of type $i - 1$. Such a point is, by definition, an element of a path $p_1p_2p_3 \cdots p_{i-1} \equiv P$. Now the vertices of P^* are all of type greater than $i - 1$ and hence by the inductive assumption are all of degree less than or equal to O_i. Since the number of components of P^* is less than or equal to the number of components of $G - P$, and this in turn is, by Theorem 3.1, less than or equal to l^l it follows from Corollary 3.2.1 that the order of P^* cannot exceed $l^l(O_i + O_i^2 + \cdots + O_i^l) + l = O_{i-1}(l, k)$. Thus the degree of a point of type $i - 1$ is a bounded function of l and k. It follows that if l and k are fixed then there is a finite upper bound to the degree of the vertices of G. Finally, since a k-critical graph is connected we conclude from Theorem 3.2 that the order of G is a bounded function of l and k or, what is the same thing, l must go to infinity with n. Now Dirac [1] has shown that a graph

without a cut node which contains a path of length $l > 1$ contains a circuit of length exceeding $(2l)^{\frac{1}{2}}$ from which fact our theorem follows.

4. Four chromatic graphs.

The following lemma asserts, in essence, that it is possible to color any circuit C in three colors subject even to the restriction that each node be denied one of the colors (the forbidden color can vary from node to node) provided that the excluded color is not the same for every node. The application of the lemma will be to 3-critical graphs which are easily seen to be odd circuits.

LEMMA 4.1. *Associated with each non-constant integer-valued function* $F(x)$, $1 \leqq F(x) \leqq 3$ *defined on the nodes of a circuit* C *there is a 3-coloring,* $f(x)$, *of* C *such that* $F(x) \neq f(x)$ *for any* $x \in C$.

Proof. Choose the labelling $a_1a_2a_3 \cdots a_n$ of the vertices of C such that $F(a_1) \neq F(a_2)$. Now define $f(a_2) = F(a_1)$ and $f(a_i) = x$, x one of the integers 1, 2, 3, but $x \neq f(a_{i-1})$ and $x \neq F(a_i)$, $i = 3, 4, \cdots, n$. Finally define $f(a_1) = x, x \neq f(a_n)$ and $x \neq F(a_1) = f(a_2)$.

THEOREM 4.1. *For infinitely many values of* n,

$$L_4(n) \leqq 3 \log^2 n / \log^2 (27/4).$$

Proof. Let T be a set of $3h - 2$ pairwise disjoint nodes, where h is an odd number. Let $\{H_r\}$ be the collection of pairwise different subsets of h elements of T, and $\{C_r\}$ a collection of replicas of C, a circuit with h nodes, $r = 1, 2, \cdots, \binom{3h-2}{h}$, and connect each node of H_r with each node of C_r in a 1-1 fashion. Let G be the graph whose nodes are the elements of T and the nodes of the various replicas of C with the edges as indicated above. Then we claim (1) G is 4-chromatic and (2) G is critical.

Proof of (1). We show first that G is not 3-chromatic. If G were 3-chromatic then some one of the colors would have to be used at least h times in coloring T, since T contains $3h - 2$ nodes. Call this color A_1 and suppose H_1 is a set of h nodes of T carrying the color A_1 in some 3-coloring of G. The replica C_1 each of whose nodes is joined with a node of H_1, is 3-chromatic since h is odd, but A_1 cannot be used in its coloring. Hence three additional colors must be used which contradicts our assumption that G is 3-chromatic.

To color G in four colors we have merely to use a single color to color the nodes of T and three additional colors to color the various replicas of C.

Proof of (2). The nodes of G are of two types: (i) those belonging to T, and (ii) those belonging to some C_i. We show first that the removal of a node of type (i) renders G 3-chromatic. We color the remaining $3h-3$ nodes of T with three colors using each color on exactly $h-1$ nodes. Then each of the C_i may be colored with the same three colors by virtue of Lemma 4.1, for in no case is the same color excluded from all h nodes of any replica of C.

Now we show that the removal of a node of type (ii) renders G 3-chromatic. There is no loss of generality in assuming that the removed node x is an element of C_1. We color the nodes of H_1 with color A_1 and distribute colors A_2 and A_3 equally among the remaining $2h-2$ nodes of T. C_1-x may now be colored with A_2 and A_3 while, in view of Lemma 4.1, the remaining replicas may be colored with the three colors. This completes the proof of 2.

We now examine the maximal circuit length in G_h. In order to pass from one replica of C to another via a circuit it is necessary to pass through one of the $3h-2$ nodes in T. It follows that the length of any circuit in G is at most $(3h-2)(h+1)$. The number of nodes in G is $\binom{3h-2}{h}h+3h-2$. Our theorem now follows from Stirling's formula.

Remark. If we could prove the analogue of Lemma 4.1 for critical k-chromatic graphs with $k>3$, then we could easily show that

$$\liminf L_k(n)/\log^2 n < \infty.$$

It would be necessary merely to parallel the steps of our proof of Theorem 4.1 replacing C by a k-critical graph satisfying the conditions of the theorem, modifying T appropriately, and using induction on k.

Dirac [1] conjectured that it might be true that every 4-critical graph is Hamiltonian (see definition 2.10). Theorem 4.1 shows that this conjecture is incorrect.

5. Minimal Circuit Length.

THEOREM 5.1. *For any k there exists a k-chromatic graph with no circuits of length $\leqq 5$.*

Proof. We use induction on k; the theorem is obvious when $k=1$. Let G_k be a k-chromatic graph with no circuits of length less than or equal to 5. Using G_k we shall construct a $k+1$ chromatic graph G_{k+1}, with no circuits of length less than or equal to 5.

Suppose G_k has α nodes. Let T be a set of $k\alpha$ nodes, pairwise disjoint. Let $\{H_r\}$, $r = 1, 2, \cdots, \binom{k\alpha}{\alpha}$ be the collection of pairwise different subsets of nodes of T. Form $\binom{k\alpha}{\alpha}$ replicas $\{G_k{}^r\}$ of G_k and connect the node of H_r with the nodes of $G_k{}^r$ in a 1-1 fashion. Let G'_{k+1} be the graph consisting of T and the replicas of G_k, the various nodes being joined as described above. Then we claim: (1) G'_{k+1} is at least $k+1$ chromatic, and (2) G'_{k+1} contains no circuits of length less than or equal to 5.

Proof of (1). Suppose G'_{k+1} were k-chromatic. Consider any coloring of G'_{k+1} by k colors. Then at most k colors are used to color T, in this coloring. Since T has $k\alpha$ nodes some color, say A, must be used at least α times in coloring T. Let H_r be a set of α nodes of T all of which are colored by A. Since $G_k{}^r$ is k chromatic and since A cannot be used in its coloring it follows that at least k colors in addition to A must be used to color G'_{k+1}. Thus G'_{k+1} is at least $k+1$ chromatic.

Proof of (2). By our inductive hypothesis, no circuit of length less than or equal to 5 in G_{k+1} can be composed only of nodes of a single replica of G_k. Such a circuit must therefore contain a node of T, since no two nodes in different replicas are joined. It is clear that the shortest possible circuit starting and ending at some node, a, of T must be constructed as follows: The circuit starts at a, then it proceeds to a point in some replica of G_k. The next node in the circuit must be in the same replica; then must come a node in T, say b, different from a. From b the circuit enters a second replica, passes through two nodes and then returns to a. This shortest possible circuit is of length 6.

If G'_{k+1} is not $k+1$ chromatic it has a $k+1$ chromatic subgraph which of course has no circuits of length less than or equal to 5. This subgraph is the desired G_{k+1}. We may suppose if we wish that G_{k+1} is critical.

Added in the proof. It has recently come to our attention that the result of Theorem 5. 1 has been obtained by W. T. Tutte. Cf. *American Mathematical Monthly*, vol. 61 (May, 1954), p. 352.

MICHIGAN STATE COLLEGE.

BIBLIOGRAPHY.

[1] G. A. Dirac, "Some theorems on abstract graphs," *Proceedings of the London Mathematical Society*, third series, vol. 11, No. 5, March, 1952.

CONCERNING THE IMBEDDING OF UPPER SEMICONTINUOUS COLLECTIONS OF CONTINUA IN CONTINUOUS COLLECTIONS OF CONTINUA.*

By MARY-ELIZABETH HAMSTROM.

1. Definitions and notation. The purpose of this paper is to establish for certain upper semicontinuous collections G of mutually exclusive continuous curves necessary and sufficient conditions that there should exist continuous collections W of mutually exclusive continuous curves such that G is imbedded in W. Throughout this paper the space considered will be the Euclidean plane, E.

Definition 1. If G and W are collections of mutually exclusive point sets, G is said to be *imbedded* in W provided it is true that each element of G is contained in some element of W and each element of W contains one and only one element of G.

If the collection G is imbedded in the collection W, then (1) if g is an element of G, $W(g)$ will denote the element of W containing g and (2) if H is a subcollection of G, $W(H)$ will denote the subcollection of W in which H is imbedded. In the remainder of this section G will denote an upper semicontinuous collection of mutually exclusive compact atriodic continuous curves which is an arc with respect to its elements. The endelements of G will be denoted by a and b.

If H is a collection of point sets the sum of the elements of H is denoted by H^*. If H is an upper semicontinuous collection of point sets a subcollection of H will be said to be closed (totally disconnected, connected, perfect, open) provided it is a closed (totally disconnected, connected, perfect, open) subset of G with respect to its elements.

Definition 2. The arc h which is an element of G is said to be a K-arc or K-curve of G provided it is true that there exist two simple closed curves J_1 and J_2 with mutually exclusive interiors I_1 and I_2 such that $J_1 \cdot J_2$ is an arc which contains and is separated by h, $J_1 \cdot J_2 \cdot G^* = h$, I_1 does not intersect any element of the interval ah of G and I_2 does not intersect any element of the interval hb. A simple closed curve which is an element of G is a

* Received October 5, 1953.

K-curve provided it either separates two points of G^* in E or is an endelement of G. A point which is an element of G is a *K-point* or *K-curve* of G provided it is not the sequential limiting set of any sequence of simple closed curves of G.

Definition 3. If n is a natural number and H is a closed subcollection of arcs of G, $H^{(n)}$ is the collection of all elements g of H such that if ϵ is a positive number then every open interval of G containing g contains an element of H which has a subarc the distance between whose endpoints is less than ϵ but whose diameter is not less than $1/n$.

Clearly $H^{(n)}$ is closed.

The terms "*link*" and "*simple chain*" will be used in the ordinary sense with the added conventions that each link of a simple chain shall be an open disc (i. e. the interior of a simple closed curve) and if x and y are links of C, $\bar{x} \cdot \bar{y}$ (if it exists) is a closed disc (i. e. the closure of an open disc) and $(\bar{x}\text{-}x) \cdot (\bar{y}\text{-}y)$ contains only two points. A simple cyclic chain is a chain, C, such that if any link is removed from C the remainder is a simple chain. The least upper bound of the diameters of the links of a chain is called its *link diameter*. If C is a simple chain containing three or more links then $[C]$ shall denote the chain obtained by omitting the endlinks from C. $[C^*]$ is the sum of the links of $[C]$, and $[\bar{C}^*]$ is the closure of the sum of the links of $[C]$. If the arc or simple closed curve, t, is covered by the simple or simple cyclic chain, C, t is said to be straight with respect to C provided that if x is a link of C intersecting t the boundary of x does not contain three points of t and if it does contain two points of t these points lie in different links of C. If t is not covered by C it is said to be straight with respect to C provided that $t \cdot \bar{C}^*$ is an arc whose cut points lie in C^*, every subarc of t lying in C^* is straight with respect to C, and if $(\bar{C}^*\text{-}C^*) \cdot t$ and the boundary of a link x of C intersect their common part is degenerate and x is an endlink of C. The simple or simple cyclic chain K is said to be straight with respect to the simple or simple cyclic chain C provided it is true that there is a subchain L of K such that (1) if x and y are links of K and C respectively then $\bar{x} \cdot \bar{y}$ is a closed disc if it exists and if x is also a link of L then $\bar{x}$ is a subset of C^*, (2) if y and z are consecutive links of C there are not two links of K intersecting z and $y\text{-}\bar{z}$, and (3) if x is a link of K not in L, $\bar{x}$ and $[\bar{C}^*]$ are mutually exclusive and if x and L^* are mutually exclusive so are $\bar{x}$ and $\bar{C}^*$. If the arc or simple closed curve t is covered by the simple or simple cyclic chain C, t is *straight* with respect to C and for no proper subchain C' of C does C'^* contain t, then C is said to *simply cover* t. (These definitions are modifications of definitions found in [2].)

2. Principal imbedding lemma.

LEMMA 1. *Let G be an upper semicontinuous collection of compact atriodic continuous curves such that* (1) *G is an arc with respect to its elements,* (2) *the endelements of G are simple closed curves of which one is interior to the other but every non-endelement of G is acyclic,* (3) *if K is a closed subcollection of arcs of G then for no natural number n is it true that $K^{(n)} = K$, and* (4) *each element of G is a K-curve of G. If H is a closed totally disconnected subcollection of G containing the endelements of G then there is a continuous collection W of simple closed curves such that H is imbedded in W, $W^* \cdot G^* = H^*$, and of each pair of elements of W, one is interior to the other.*

Proof. For the purposes of this lemma only, Definition 3 will be applied to closed subcollections K of G which contain a or b. For each natural number n there is a well-ordered sequence α_n such that (1) the first element of α_n is H, (2) if the element A of α_n has a successor the first element of α_n following A is $A^{(n)}$, (3) if α'_n is a subsequence of α_n which has no last element then the first element of α_n following every element of α'_n is the common part of the elements of α'_n, and (4) α_n has a last element, K, and $K^{(n)}$ does not exist.

If A is an element of α_n and $A^{(n)}$ exists, $A^{(n)}$ is a closed proper subcollection of A. Consequently α_n is countable. Let Γ_n denote the collection of all elements of α_n. Then the collection $\Sigma\Gamma_i$ is countable and its elements can be arranged in a simple sequence $A_1, A_2, A_3, \cdots$.

Let a and b denote the endelements of G, suppose that a is exterior to b, and let I denote the set of points interior to a and exterior to b. An *I set* shall be an interval of G which meets H in a set which is both open and closed in H.

If h is an element of H-a-b let $C_1(h)$ denote a simple chain with link diameter less than 1 simply covering h such that (1) $\bar{C}_1^*(h)$ is[1] a subset of I and (2) if h is not in A_1 then A_1^* and $\bar{C}_1^*(h)$ are mutually exclusive. If h is a or b let $C_1(h)$ denote a simple cyclic chain with link diameter less than 1 simply covering h such that if h is not in A_1 then A_1^* and $\bar{C}_1^*(h)$ are mutually exclusive.

Each element h of H is contained in an I set each of whose elements lies in $C_1^*(h)$. Consequently there exist a finite subcollection H_1 of H and a

[1] For typographical convenience, $\bar{C}_1^*(h)$ will denote the closure of $C_1^*(h)$.

finite collection S_1 of non-intersecting I sets covering H such that each element of H_1 belongs to some I set of S_1, each I set of S_1 contains one and only one element of H_1, and each continuum of the I set of S_1 containing the element h of H_1 is a subset of $C_1^*(h)$. Let $G_1(h)$ denote the I set of S_1 containing the element h of H.

Since every element of H_1 is a K-curve of G there is a collection $W(H_1)$ of mutually exclusive simple closed curves such that H_1 is imbedded in $W(H_1)$, $H_1^* = G^* \cdot W^*(H_1)$, and of each pair of elements of $W(H_1)$, one is interior to the other. If h is an element of H-a-b, let $K_1(h)$ denote a simple chain with link diameter less than 1 simply covering the closure of $W(h)$-h such that $K_1^*(h)$ is a subset of I and if g is an element of H_1-h then $[\bar{K}_1^*(h)]$ and $G^* + [\bar{K}_1^*(g)]$ are mutually exclusive. If h is a or b let $K_1(h)$ be identical with $C_1(h)$.

If h is an element of $H - H_1$, let $C_2(h)$ denote a simple chain with link diameter less than $1/2$ simply covering h such that (1) if h is not in A_j $(j = 1, 2)$ then $\bar{C}_2^*(h)$ and A^*_j are mutually exclusive and (2) if h is in the I set $G_1(g)$ of S_1 then (i) if x is a link of $C_2(h)$ there is a link of $C_1(g)$ containing $\bar{x}$ but there are not three links of $C_1(g)$ whose closures intersect $\bar{x}$ and (ii) if g' is an element of $H_1 - g$ then $\bar{C}_2^*(h)$ does not intersect $[\bar{K}_1^*(g)] + W^*(H_1) + [\bar{K}_1^*(g')] + G_1^*(g')$.

If h and h' are elements of H_1 let $L_2(h)$ be a simple cyclic chain with link diameter less than $1/2$ simply covering $W(h)$ such that (1) $L_2(h)$ is straight with respect to $[K_1(h)]$, (2) $\bar{L}_2^*(h)$ and $\bar{L}_2^*(h')$ are mutually exclusive, (3) if x is a link of $L_2(h)$ there is a link of $C_1(h)$ or $K_1(h)$ which contains $\bar{x}$ but neither $C_1(h)$ nor $K_1(h)$ contains three links whose closures intersect $\bar{x}$, (4) if h is not in A_j $(j = 1, 2)$, $\bar{L}_2^*(h)$ and A_j^* are mutually exclusive and (5) if h is in $H_1 - a$-b there is a simple subchain $C_2(h)$ of $L_2(h)$ which simply covers h and has the property that no link of $L_2(h)$ not in $C_2(h)$ has a closure which intersects G^*. Let $C_2(h)$ and $K_2(h)$ be $L_2(h)$ if h is a or b. Otherwise let $K_2(h)$ be the simple chain whose links are the links of $L_2(h)$ not in $C_2(h)$.

Each element h of H is contained in an I set each of whose elements lies in $C_2^*(h)$. Consequently there exist a finite subcollection H_2 of H and a finite collection, S_2, of non-intersecting I sets covering H such that (1) H_1 is a subcollection of H_2, (2) each element of H_2 is an element of some I set of S_2 and each element of S_2 is a subcollection of some I set of S_1 and contains one and only one element of H_2, (3) each continuum of the I set of S_2 containing the element h of H_2 is a subset of $C_2^*(h)$. Let $G_2(h)$ denote the I set of S_2 containing the element h of H_2.

There is a collection $W(H_2)$ of mutually exclusive simple closed curves which contains $W(H_1)$ such that (1) H_2 is imbedded in $W(H_2)$ and $H_2^* = G^* \cdot W^*(H_2)$, (2) if h is an element of H_2 and the I set $G_1(h')$ of S_1 then $W(h)$ is straight with respect to $[K_1(h')]$ and intersects each link of $C_1(h')$ and $K_1(h')$, and (3) of each pair of elements of $W(H_2)$, one is interior to the other. For each element h of H_1 let $N_{11}(h)$ be the number of arcs x of $W^*(H_2 \cdot G_1(h))$ for which there is a link c of $C_1(h)$ such that x is a subset of c and is irreducible between the closures of two non-intersecting links of $C_1(h)$. Let $N_{11}(W)$ be the sum of the $N_{11}(h)$ for h in H_1. Then $W(H_2)$ is to have the further property that if $W'(H_2)$ is any other collection satisfying properties (1) and (2) above then $N_{11}(W') \geqq N_{11}(W)$.

If h is an element of $H_2 - H_1$ and the I set $G_1(h')$ of S_1 let $K_2(h)$ denote a simple chain with link diameter less than $1/2$ simply covering the closure of $W(h) - h$ such that (1) if x is a link of $K_2(h)$ there is a link of $C_1(h')$ or $K_1(h')$ containing $\bar{x}$ but neither of these chains contains three links whose closures intersect $\bar{x}$, (2) $[K_2(h)]$ is straight with respect to $[K_1(h')]$ and (3) if g is an element of $H_2 - h$, $[\bar{K}_2^*(h)]$ does not intersect $W^*(H_1) + G^* + [\bar{K}_2^*(g)]$.

If we continue in this fashion it can be shown that for each natural number i there exist collections H_i, $W(H_i)$, and S_i and that for each element h of H_i there are an I set $G_i(h)$ and simple chains $C_i(h)$ and $K_i(h)$ and that these collections and chains satisfy the following conditions.

[1] H_{i+1} is a finite subcollection of H containing $H_i + a + b$ and it is imbedded in $W(H_{i+1})$, a collection of mutually exclusive simple closed curves containing $W(H_i)$ such that $W^*(H_{i+1}) \cdot G^* = H_{i+1}^*$ and of each pair of elements of $W(H_{i+1})$, one is interior to the other.

[2] If h is in H_{i+1}, $G_{i+1}(h)$ is an I set in the collection S_{i+1}, it contains h, it is subcollection of some I set of S_i, each of its elements lies in $C_{i+1}^*(h)$, and if h is not an element of A_j $(j \leqq i+1)$ then $A_j^* + H_{i+1}^* - h$ and $G_{i+1}^*(h)$ are mutually exclusive.

[3] S_{i+1} is the collection of all $G_{i+1}(h)$ for h is H_{i+1}, it covers H, it is finite, and its elements are mutually exclusive.

[4] If h is an element of $H_{i+1} - H_i$, $C_{i+1}(h)$ denotes a simple chain with link diameter less than $1/i+1$ simply covering h such that if h is not in A_j $(j \leqq i+1)$ then $\bar{C}_{i+1}^*(h)$ and A_j^* are mutually exclusive and if h is in the I set $G_i(g)$ of S_i then (i) if x is a link of $C_{i+1}(h)$ there is a link of $C_i(g)$ containing $\bar{x}$ but there are not three links of $C_i(g)$ whose closures intersect $\bar{x}$ and (ii) if g' is an element of $H_i - g$, $\bar{C}_{i+1}^*(h)$ does not intersect $[\bar{K}_i^*(g)] + W^*(H_i) + G_i^*(g') + [\bar{K}_i^*(g')]$.

[5] If h and h' are elements of H_i, $L_{i+1}(h)$ is a simple cyclic chain with link diameter less than $1/i+1$ simply covering $W(h)$ such that (i) $L_{i+1}(h)$ is straight with respect to $[K_i(h)]$, (ii) $\bar{L}_{i+1}{}^*(h)$ and $\bar{L}_{i+1}{}^*(h')$ are mutually exclusive, (iii) if x is a link of $L_{i+1}(h)$ there is a link of $C_i(h)$ or $K_i(h)$ containing $\bar{x}$ but neither of these chains has three links whose closures intersect $\bar{x}$, (iv) if h is not in A_j $(j \leqq i+1)$ then $A_j{}^*$ and $\bar{L}_{i+1}{}^*(h)$ are mutually exclusive and (v) if h is not a or b there is a simple subchain $C_{i+1}(h)$ of $L_{i+1}(h)$ which simply covers h and has the property that no link of $L_{i+1},h)$ not in $C_{i+1}(h)$ has a closure which intersects G^*. Let $C_{i+1}(h)$ and $K_{i+1}(h)$ be $L_{i+1}(h)$ if h is a or b; otherwise let $K_{i+1}(h)$ denote the simple chain whose links are the links of $L_{i+1}(h)$ not in $C_{i+1}(h)$.

[6] If h is in $(H_{i+1}-H_i)\cdot G_j(h_j)$, $(j \leqq i)$, $W(h)$ is straight with respect to $[K_j(h_j)]$ and intersects each link of $C_j(h_j)$ and $K_j(h_j)$. For each element h of H_j let $N_{ij}(h)$ be the number of arcs x of $W^*(H_{i+1}\cdot G_j(h))$ for which there exists a link c of $C_j(h)$ such that x is a subset of c and is irreducible between the closures of two non-intersecting links of $C_j(h)$ and let $N_{ij}(W)$ be the sum of the $N_{ij}(h)$ for h in H_j. The collection $W(H_{i+1})$ is to have the further property that if $W'(H_{i+1})$ is any other collection having properties [1] and [6] then $N_{i1}(W) \leqq N_{i1}(W')$ and if $N_{ij}(W) = N_{ij}(W')$ then $N_{ij+1}(W) \leqq N_{ij+1}(W')$.

[7] If h is an element of $H_{i+1}-H_i$ and the I set $G_i(h')$ of S_i, $K_{i+1}(h)$ denotes a simple chain with link diameter less than $1/i+1$ simply covering the closure of $W(h)-h$ such that (i) if x is a link of $K_{i+1}(h)$ there is a link of $C_i(h')$ or $K_i(h')$ containing $\bar{x}$ but neither of these chains has three links whose closures intersect $\bar{x}$, (ii) $[K_{i+1}(h)]$ is straight with respect to $[K_i(h')]$ and (iii) if g is an element of $H_{i+1}-h$, $[\bar{K}_{i+1}{}^*(h)]$ does not intersect $W^*(H_i)+G^*+[\bar{K}_{i+1}{}^*(g)]$.

If h is an element of H there is only one sequence $G(h)$, $(g_1, g_2, g_3, \cdots)$ such that for each i, g_{i+1} is an I set of S_{i+1} containing h and is a subcollection of g_i. If, for each i, $g_i = G_i(h_i)$ let k_i denote $K_i{}^*(h_i)+C_i{}^*(h_i)$ and let $K(h)$ denote the sequence $k_1, k_2, k_3, \cdots$. Since for each i, $\bar{k}_{i+1}+h$ is a subset of k_i, h is a subset of the common part of the elements of $K(h)$. If h is an element of ΣH_i, $W(h)$ is the common part of the elements of $K(h)$; if h is an element of $H-\Sigma H_i$, let $W(h)$ denote the common part of the elements of $K(h)$. It readily follows from the construction that the collection W of all $W(h)$'s is a continuous collection of mutually exclusive continua no one of which intersects two elements of G. It remains to be shown that every element of W is a simple closed curve.

Suppose h is an element of $H-\Sigma H_i$ and that for some natural number

n and each positive number ϵ there is a natural number $n(\epsilon)$ such that $h_{n(\epsilon)}$ contains an arc of diameter not less than $1/n$ but the distance between whose endpoints is less than ϵ. The arc h is thus an element of $H^{(n)}$. Suppose that there is an element of α_n which does not contain h. Let A be the first such element of α_n. Clearly A has an immediate predecessor, B, B contains h, and $B^{(n)} = A$. The collection B is an element of the sequence $A_1, A_2, \cdots$; suppose $B = A_i$. If h_j $(j \geqq i)$ is not an element of B, g_j does not intersect B. Since g_j and B contain h this is impossible. Therefore if $j \geqq i$, B contains h_j; hence h is an element of $B^{(n)}$. Consequently every element of α_n contains h. In particular, the last element, K, of α_n contains h and a repetition of the above argument shows that all but a finite number of the h_i's belong to K. This implies that h lies in $K^{(n)}$, which is impossible since there is no $K^{(n)}$. Thus, if n is a natural number there is a positive number ϵ such that for each i, if h_i contains an arc the distance between whose endpoints is less than ϵ then the diameter of that arc is less than $1/n$.

Suppose h is an element of $H - \Sigma H_i$. It follows from the construction of the chains $[K_n(h_n)]$ and the methods used in the proof of Theorem 1, Chapter II of [7] that $W(h) - h$ is topologically equivalent to an arc minus its endpoints. Consequently, if $W(h)$ contains a point at which it is not locally connected that point is in h.

Since h is a K-arc there are simple closed curves J_1 and J_2 with interiors I_1 and I_2 such that $J_1 \cdot J_2$ is an arc containing and separated by h, $J_1 \cdot J_2 \cdot G^* = h$, I_1 does not intersect any element of the interval ah of G and I_2 does not intersect any element of hb.

If $W(h)$ is not locally connected there exist a subarc x of h and a sequence of mutually exclusive subarcs of $W(h) - h$ converging to x. Let δ denote the diameter of x. It follows from the arguments of the third paragraph above that there is a natural number N such that $1/N < \delta/100$ and (1) for each j, if h_j contains an arc the distance between whose endpoints is less than $1/N$ then that arc has diameter less than $\delta/100$ and (2) $\bar{C}_N^*(h_N)$ is a subset of $I_1 + I_2 + J_1 \cdot J_2$ minus its boundary. There exist a subarc y of x the distance between whose endpoints is δ, a sequence β of mutually exclusive subarcs of $W(h) - h$ converging to y, and a minimal subchain C of $C_N(h_N)$ covering y such that $\beta^* + y$ is a subset of C^*. If A and B are the endlinks of C the distance between them exceeds $.98\delta$. Let D_1, D_2, and D_3 be consecutive links of C in that order no one of which intersects $A + B$ and let D denote $D_2 - (D_1 + D_3) \cdot D_2$.

It follows from the definition of N that if $j > N$ there do not exist two subarcs of h_j both irreducible between A and B. Therefore, since the sequential

limiting set of $W(h_1), W(h_2), \cdots$ contains all the elements of β and all but a finite number of the elements of β intersect both A and B there exist for each natural number q and each positive number ϵ a neighborhood U of h lying in an ϵ-neighborhood of h and a natural number M such that if $k \geqq M$, h_k lies in U and $W(h_k) - h_k$ contains q arcs each lying in $C^* - C^* \cdot \bar{U}$ and irreducible between $\bar{A}$ and $\bar{B}$. In particular, there are a natural number p and a collection λ of three subarcs of $W(h_p) - h_p$ such that (1) each element of λ is a subset of D irreducible between $\bar{D}_1$ and $\bar{D}_3$, (2) λ^* lies either in I_1 or in I_2, (3) no two elements of λ are separated in D by $h + J_1 + J_2$, any element of $W(H_N)$ or any element of $W(H_j)$, $(j \leqq p)$, not in $G_N(h_N)$ and (4) no component of $D - \lambda^*$ whose boundary contains two arcs of λ intersects $\bar{C}_p^*(h_p)$.

The arcs of λ can be denoted by u_1, u_2, and u_3 in such a way that (1) in D, u_2 separates u_1 minus its endpoints from u_3 minus its endpoints and (2) if E_1 and E_2 denote the interior and exterior of $W(h_p)$ respectively then for $i = 1, 2$ there is an arc t_i lying in D irreducible between u_i and u_{i+1} and lying except for its endpoints in E_i.

Suppose that h is exterior to $W(h_p)$ and λ^* lies in I_1; since u_2 and u_3 lie in I_1 and are not separated by $J_1 + J_2$ in D, the sum of the elements of the interval ah of G does not separate u_2 from u_3 in D. Since t_2 lies except for its endpoints in the exterior of $W(h_p)$ the sum of the elements of the interval $h_p b$ does not separate u_2 from u_3 in D. Since t_2 does not intersect $C_p^*(h_p)$ and the elements of the interval hh_p all lie in $C_p^*(h_p)$, u_2 and u_3 are not separated in D by the sum of the elements of hh_p. No element of $W(H_N)$ separates u_2 from u_3 in D. Suppose there is a natural number m between N and $p+1$ such that some element of $W(H_m)$ separates u_2 from u_3 in D or contains $u_2 + u_3$, but no element of $W(H_{m-1})$ separates u_2 from u_3 in D. There is an element w of $W(H_m)$ containing two arcs w_1 and w_2 lying in D such that each is irreducible between $\bar{D}_1$ and $\bar{D}_2$, neither lies entirely in G^*, each separates u_2 from u_3 in D and no element of $W(H_{m-1})$ separates w_1 from w_2. It follows from (3) above that w belongs to $W(G_N(H_N))$. There is an arc t lying in D which is irreducible between w_1 and w_2 and does not intersect G^* or any element of $W(H_m)$, other than w. This contradicts condition [6] required of $W(H_m)$ since the existence of t clearly shows that $N_{mN}(W)$ is not a minimum.

If h is interior to $W(h_p)$ and λ^* lies in I_1 then the sum of the elements of the interval ah does not separate u_2 from u_3 in D and since the elements of the interval hb are interior to $W(h_p)$, the sum of these elements does not separate u_2 from u_3 in D. It follows from an argument similar to the above that this leads to a contradiction, as does the assumption that λ^* lies in I_2.

Thus, in any case, the assumption that $W(h)$ is not locally connected leads to a contradiction. It follows from the construction of $C_i(h_i)$ and $K_i(h_i)$ and the argument used to show that $W(h) - h$ is topologically equivalent to an arc minus its endpoints that $W(h)$ contains no triod and is a simple closed curve. The proof of Lemma 1 is now complete.

3. Lemmas concerning upper semicontinuous collections.

LEMMA 2. *Suppose the upper semicontinuous collection G of mutually exclusive continua fills up the compact continuum M and is a continuous curve with respect to its elements. If D is an open disc (the interior of a simple closed curve) and $N_1, N_2, N_3, \cdots$ is a sequence of distinct components of $M \cdot \bar{D}$ converging to a continuum N which intersects D but no element of $N_1, N_2, N_3, \cdots$, then N is a subset of some element of G.*

Proof. Suppose the lemma is false. With the aid of Theorem 8, Chapter II of [7] and some arguments used in the proof of Theorem 2 of [6] it can be shown that there are a subcontinuum K of N intersecting two elements of G, a simple closed curve J whose interior, I, intersects K, and a sequence $K_1, K_2, K_3, \cdots$ of distinct components of $M \cdot \bar{I}$ converging to K such that for each natural number i, K_{i+1} separates K_i from K in $M \cdot \bar{I}$. The sequential limiting set, L, of the sequence $J \cdot K_1, J \cdot K_2, \cdots$ does not contain three points. Consequently there is an element x of G which intersects $K \cdot I$ but not L. There is a connected open subcollection U of G containing x such that $\bar{U}^*$ and L do not intersect. Since U^* is open and intersects K it intersects all but a finite number of the K_i and since U^* is connected and the K_i are components of $M \cdot \bar{I}$, it intersects all but a finite number of the elements of $K_1 \cdot J, K_2 \cdot J, \cdots$. This implies that $\bar{U}^*$ intersects L. Thus the supposition that Lemma 2 is false leads to a contradiction.

LEMMA 3. *If G is an upper semicontinuous collection of mutually exclusive arcs which is an arc with respect to its elements then if G is continuous*[2] *at an element g which contains a point at which G^* is not locally connected then there is a natural number n such that g belongs to $G^{(n)}$.*

Proof. It follows from Lemma 2, the proof of Theorem 2 of [6] and Theorem 8, Chapter II, of [7] that there is an open disc D intersecting g such that $g \cdot \bar{D}$ is an arc and that there is a sequence α of mutually exclusive

[2] The upper semicontinuous collection G of compact closed point sets filling up a compact closed point set is continuous at the element g provided that if a sequence of elements of G converges to a subset of g it converges to g.

components of $\bar{D} \cdot G_i^*$ converging to $g \cdot \bar{D}$ none of whose elements intersects g. Let P be a point of $g \cdot D$ and let n be a natural number such that the distance from P to $\bar{D} - D$ exceeds $1/n$. If ϵ is a positive number let U be a domain containing P such that $\bar{U}$ is a subset of D with diameter less than ϵ and the distance from $\bar{U}$ to $\bar{D} - D$ exceeds $1/n$. Since G is continuous at g there is an interval H of G containing g as an endlement such that H^* contains infinitely many elements of α and every element of H intersects U. Let k be an element of α which intersects U and lies in H^*. Since no connected subset of $H^* \cdot \bar{D}$ intersects both k and g, $H^* \cdot \bar{D}$ is the sum of two mutually separated point sets L and N, L containing g and N containing k. The point set $\bar{U} \cdot H^*$ is the sum of the mutually separated point sets $L \cdot \bar{U} \cdot H^*$ and $N \cdot \bar{U} \cdot H^*$. Since H^* is connected there is an element h of H which intersects both $L \cdot \bar{U} \cdot H^*$ and $N \cdot \bar{U} \cdot H^*$. The subarc h' of h irreducible between $L \cdot \bar{U} \cdot H^*$ and $N \cdot \bar{U} \cdot H^*$ is not a subset of $\bar{D}$ since $L \cdot H^* \cdot \bar{D}$ and $N \cdot H^* \cdot \bar{D}$ are mutually separated. Thus h' intersects $\bar{D} - D$ and U and consequently its diameter exceeds $1/n$. Since its endpoints lie in U the distance between them is less than ϵ. Since every segment of G containing g contains such an element, h, g belongs to $G^{(n)}$.

LEMMA 4. *If G satisfies the conditions required of it in Lemma* 1 *then if for the natural number n, K_n is the set of all elements x of G for which there is a sequence $\alpha(x)$ of elements of G converging to a subset y of x such that $x - y$ has a component whose diameter exceeds $1/n$ then K_n is nowhere dense in G.*

Proof. Suppose there is an interval V of acyclic elements of G in which K_n is dense. Since for each natural number i, $G^{(i)}$ is closed and totally disconnected there is an element x_1 of $K_n \cdot V$ which is not in $G^{(1)}$ and there is a subinterval V_1 of V containing x_1 as a non-endelement but containing no element of $G^{(1)}$. Let y_1 be the sequential limiting set of $\alpha(x_1)$. There is a domain D_1 containing y_1 each point of which is closer to y_1 than 1 and such that there is a component z_1 of $x_1 - \bar{D}_1 \cdot x_1$ whose diameter exceeds $1/n$. Since G is an upper semicontinuous collection and D contains all but a finite number of the elements of $\alpha(x_1)$ there is an element x_2 of $K_n \cdot V_1$ which lies in D_1 but is not an element of $G^{(2)}$ and there is a subinterval V_2 of V_1 containing x_2 as a non-endelement such that each of its elements lies in D_1 but none of them belongs to $G^{(2)}$. Let y_2 be the sequential limiting set of $\alpha(x_2)$. There is a domain D_2 containing y_2 each point of which is closer to y_2 than 1/2 and such that $\bar{D}_2$ is a subset of D_1 and there is a component z_2 of $x_2 - \bar{D}_2 \cdot x_2$ whose diameter exceeds $1/n$.

In general, for each natural number i there are continua x_i, y_i, and z_i, a domain D_i, and an interval V_i of V such that (1) x_i is an element of $K_n \cdot V_i$ but not an endelement of V_i, (2) V_i and $G^{(i)}$ are mutually exclusive and V_{i+1} is a subset of V_i, (3) $V_{i+1}{}^*$ is a subset of D_i, (4) y_i is the sequential limiting set of $\alpha(x_i)$ and D_{i+1} is a domain containing y_{i+1} each point of which is closer to y_{i+1} than $1/i+1$ and such that $\bar{D}_{i+1}$ lies in D_i and there is a component z_{i+1} of $x_{i+1} - \bar{D}_{i+1} \cdot x_{i+1}$ whose diameter exceeds $1/n$.

Let z be the sequential limiting set of an infinite subsequence $z_1', z_2', \cdots$ of $z_1, z_2, \cdots$, and suppose z is a subset of the element x of G. Since for each i, x is in V_i, x belongs to $G - \Sigma G^{(i)}$. For each i, x_j, $(j > i)$, and consequently z_j lies in D_i and thus each point of z_j is within distance $1/i$ of y_i. Therefore if P and Q are points of z and D_P and D_Q are domains containing P and Q respectively such that $\bar{D}_P$ and $\bar{D}_Q$ are mutually exclusive there is a natural number N such that if $j > N$, D_P and D_Q both intersect the mutually exclusive subcontinua y_j' and z_j' of the arc x_j'. This is sufficient to show that for some natural number m, x is in $G^{(m)}$, which is impossible. Thus the supposition that K_n is not nowhere dense leads to a contradiction.

LEMMA 5. *If W is a continuous collection of mutually exclusive simple closed curves filling up a compact continuous curve M such that W is an arc with respect to its elements and H is an upper semicontinuous collection*[3] *of mutually exclusive arcs which is imbedded in a closed subcollection of W then for no positive integer n is it true that $H^{(n)} = H$.*

Proof. If a and c are elements of H the notation $I(ac)$ will be used to denote the collection consisting of a, c, and all elements b of H such that $W(b)$ is between $W(a)$ and $W(c)$. Suppose that for some natural number n, $H^{(n)} = H$. Since every element of H is thus a limit element of H, $W(H)$ is a perfect subcollection of W.

Supopse that the element x of H is a limit element of $I(x, y)$ and that there is a finite sequence $x_1, x_2, \cdots, x_p$ of mutually exclusive subarcs of $W(x)$ each having diameter greater than $1/4n$. There is a simple cyclic chain, C, simply covering $W(x)$ such that if i and j are integers not exceeding p and s_i and s_j are links of C intersecting x_i and x_j respectively then $\bar{s}_i$ and $\bar{s}_j$ are mutually exclusive. For each i let C_i denote the minimal subchain of C simply covering x_i. Since x is in $H^{(n)}$ there is a subarc x' of x whose diameter exceeds $1/4n$ such that (1) x' does not intersect three arcs of the sequence $x_1, x_2, \cdots, x_p$ and (2) if K is the simple subchain of C simply covering x'

[3] If, in Definition 3, the words "open interval" be replaced by "open set" the definition of H will apply here.

then if z is an element of $I(x, y) - x$ there is an element of $I(x, z)$ containing two mutually exclusive arcs of diameter exceeding $1/4n$ which lie in K^* and intersect both endlinks of K. Since W is a continuous collection of simple closed curves it follows from [1] and [3] that W^* is an annulus whose boundary simple closed curves are the endelements of W. Therefore there is an element z of $I(x, y)$ such that (1) z is a limit element of $I(x, z)$, (2) z contains two mutually exclusive arcs of diameter exceeding $1/4n$ which lie in K and intersect both endlinks of K and (3) if z' is an element of $I(x, z)$ then $W(z')$ contains the elements of a sequence $z_1, z_2, \cdots, z_p$ of mutually exclusive arcs such that for each i, z_i lies in C_i^* and intersects both endlinks of C_i. Since z itself has property (3), property (2) and the fact that $W(z)$ is a simple closed curve can be used to show that $W(z)$ contains $p+1$ mutually exclusive arcs with diameter exceeding $1/4n$. Since z is a limit element of $I(x, z)$ the above argument can be repeated.

Since each element of H is in $H^{(n)}$ there is an element x_0 of H containing two mutually exclusive arcs of diameter exceeding $1/4n$. It follows from this and the argument of the above paragraph that there is a sequence $I_1, I_2, \cdots$ of intervals of W such that for each i, I_{i+1} is a subinterval of I_i and each element of I_i contains i mutually exclusive arcs of diameter exceeding $1/4n$. There is an element w of W lying in the common part of the I_i. Consequently, for each i, w contains i mutually exclusive arcs of diameter exceeding $1/4n$. This is impossible, since w is a simple closed curve. Hence the supposition that $H = H^{(n)}$ leads to a contradiction.

4. Imbedding theorems.

THEOREM 1. *Under the conditions of Lemma* 1 *there is a continuous collection* W *of simple closed curves such that* G *is imbedded in* W.

Indication of proof. Let H denote $\Sigma G^{(i)} + \Sigma \bar{K}_i$. It follows from Lemmas 3 and 4 that if g is an element of G-H, G is continuous at g and G^* is locally connected at every point of g. Since, for each i, $G^{(i)}$ and $\bar{K}_i$ are closed and totally disconnected, H is the sum of the elements of a countable collection H' of mutually exclusive closed and totally disconnected subcollections of G. Thus there is a sequence $\Delta: H_1, H_2, \cdots$ of mutually exclusive closed and totally disconnected subcollections of G the sum of whose elements is dense in G and such that every element of H' is an element of Δ and every element of Δ not in H' consists of only one element of G. Suppose that H_1 contains the endelements of G.

There is a continuous collection W_1 of simple closed curves such that

H_1 is imbedded in W_1, $W_1^* \cdot G^* = H_1^*$, and of each pair of elements of W_1, one is interior to the other. Suppose U is a component of $I - I \cdot W_1^*$ and is bounded by the simple closed curves w and w' (where I is as defined in the proof of Lemma 1). Let u_1 and u_2 be mutually exclusive arcs lying in $I - G^*$ and irreducible between w and w' such that if P is a point of the component of $U - U \cdot (u_1 + u_2)$ containing $G^* \cdot U$ then the distance from P to $G^* \cdot U$ is less than 1. Let V be the component of $U - U \cdot (u_1 + u_2)$ which does not intersect G^*. There is a continuous collection Z of mutually exclusive arcs which fills up V such that (1) Z is an arc with respect to its elements, (2) the endelements, u and u', of Z lie on w and w' respectively and (3) each arc which is an element of Z is irreducible between u_1 and u_2.

If g and g' are the elements of G imbedded in w and w' respectively there is a reversibly continuous transformation, T, of the interval gg' of G onto Z such that $T(g) = u$ and $T(g') = u'$. Let Γ be the common part of the interval gg' and the collection $H_1 + H_2$. There is a continuous collection W_U of mutually exclusive simple closed curves such that (1) Γ and $T(\Gamma)$ are imbedded in W_U, (2) $W_U^* \cdot G^* = \Gamma^*$ and $W_U^* \cdot Z^* = [T(\Gamma)]^*$, (3) w and w' belong to W_U, and (4) if h is an element of Γ, h and $T(h)$ are imbedded in the same element of W_U. Let W_2 be the collection such that h is an element of it if and only if either there is a compoent U of $I - I \cdot W_1^*$ such that h belongs to W_U or h belongs to W_1.

Suppose U' is a component of $U - U \cdot W_2^*$ and is bounded by the simple closed curves w_1 and w_1'. The point set $\bar{U}' \cdot u_1$ and $\bar{U}' \cdot u_2$ are arcs irreducible between w_1 and w_1'. Let u_1' and u_2' be mutually exclusive arcs lying in $\bar{U}' - \bar{U}' \cdot (G^* + \bar{V})$ irreducible between w_1 and w_1' such that (1) if P is a point of the component of $U' - U' \cdot (u_1' + u_2')$ containing $G^* \cdot U'$ then the distance from P to $G^* \cdot U'$ is less than 1/2 and (2) $u_1' + u_2'$ separates $\bar{U}' \cdot \bar{V}$ from $G^* \cdot \bar{U}'$ in $\bar{U}'$. Let V' be the component of $U' - U' \cdot (u_1' + u_2')$ which contains $\bar{U}' \cdot \bar{V}$. There is a continuous collection Z' of mutually exclusive arcs which fills up $\bar{V}'$ such that (1) Z' is an arc with respect to its elements, (2) the endelements, v and v', of Z lie on w_1 and w_1', respectively, (3) each arc of Z' is irreducible between u_1' and u_2' and (4) the collection of elements of Z lying in $\bar{V}'$ is imbedded in Z'.

If g_1 and g_1' are the elements of G imbedded in w_1 and w_1' there is a reversibly continuous transformation T' of the interval $g_1 g_1'$ of G into Z' such that $T'(g_1) = v$, $T'(g_1') = v'$ and if h is in the interval $g_1 g_1'$, $T(h)$ is imbedded in $T'(h)$. Let Γ_1 be the common part of $g_1 g_1'$ and $H_1 + H_2 + H_3$. There is a continuous collection $W_{U'}$ of mutually exclusive simple closed curves such that (1) Γ_1 and $T'(\Gamma_1)$ are imbedded in $W_{U'}$ (2) $W_{U'}^* \cdot G^* = \Gamma_1^*$

and $W_{U'}^* \cdot Z'^* = [T'(\Gamma_1)]^*$, (3) w_1 and w_1' belong to $W_{U'}$, and (4) if h is an element of Γ, h and $T(h)$ are imbedded in the same element of $W_{U'}$. Let W_3 be the collection such that h is an element of it if and only if either there is a component U' of $I - I \cdot W_2^*$ such that h belongs to $W_{U'}$ or h belongs to W_2.

A continuation of the process indicated above will yield the desired collection W, which consists of ΣW_i plus its limit elements. That if h is an element of $G - \Sigma H_i$ then $W(h)$ is a simple closed curve follows from the construction of Z, Z', etc. and the fact that G is continuous at h and G^* is locally connected at h.

THEOREM 2. *If G is an upper semicontinuous collection of mutually exclusive atriodic continuous curves which is an arc with respect to its elements then there exists a continuous collection W of mutually exclusive simple closed curves such that G is imbedded in W if and only if* (1) *each element of G is a K-curve of G,* (2) *if H is a closed subcollection of arcs of G then for no natural number n is it true that $H^{(n)} = H$ and* (3) *of each pair of simple closed curves which are elements of G, one is interior to the other.*

Proof. Suppose there is such a collection, W. It follows from Anderson's result, [1] and Cohen's result, [3], which state that W^* is an annulus whose boundary simple closed curves are the endelements of W that conditions (1) and (3) hold true. The truth of condition (2) follows from Lemma 5.

Suppose G satisfies conditions (1), (2), and (3). It follows from (1) and (3) that the collection, H, of all simple closed curves of G is closed and continuous. If g and g' are simple closed curves which are elements of G such that no element of G between them is a simple closed curve, the elements of the interval gg' satisfy the conditions of Theorem 1. That the conditions of 2 are sufficient follows from this and the fact that H is continuous.

LEMMA 6. *If G is an upper semicontinuous collection of compact atriodic continuous curves such that* (1) *G is an arc with respect to its elements,* (2) *the endelements of G are simple closed curves J_1 and J_2 with exteriors E_1 and E_2 such that J_1 and J_2 lie in E_2 and E_1 respectively and contain points P_1 and P_2 which are respectively accessible from the same point of $E_1 \cdot E_2 - E_1 \cdot E_2 \cdot G^*$,* (3) *if K is a closed subcollection of arcs of G then for no natural number n is it true that $K^{(n)} = K$ and* (4) *each element of G is a K-curve of G, then there is a continuous collection W of simple closed curves and arcs such that G is imbedded in W and W^* is a continuous curve.*

Proof. It follows from (2) and (4) that every non-endelement of G is

acyclic and that there is a simple closed curve J whose interior I contains $G^* - (J_1 + J_2) \cdot J$ such that $J \cdot J_1$ and $J \cdot J_2$ are points P_1 and P_2 respectively. The curve J is the sum of two non-overlapping arcs P_1XP_2 and $P_1X'P_2$. There are reversibly continuous transformations T and T' throwing G onto P_1XP_2 and $P_1X'P_2$ respectively such that $T(J_1) = P_1 = T'(J_1)$ and $T(J_2) = P_2 = T'(J_2)$. The methods used in the proofs of Lemma 1 and Theorem 1 may now be used to construct the collection W, where $W^* = J + I$ minus the sum of the interiors of J_1 and J_2 and if h is a non-endelement of G then $W(h)$ is an arc with endpoints $T(h)$ and $T'(h)$.

THEOREM 3. *If G is an upper semicontinuous collection of compact atriodic continuous curves which is an arc with respect to its elements then there is a continuous collection W of mutually exclusive compact atriodic continuous curves such that G is imbedded in W and W^* is a continuous curve if and only if every non-endelement of G is a K-curve of G and if H is a closed subcollection of arcs of G then for no natural number n is it true that $H^{(n)} = H$.*

Proof. Suppose G is imbedded in a collection W which satisfies the required conditions. If the non-endelement g of G is imbedded in the simple closed curve $W(g)$ of W, it follows from the methods used in the proof of Lemma 1 of [4] that since W^* is a continuous curve, $W(g)$ separates two points of W^* in E. Thus g is a K-curve of G. It follows from this property of the simple closed curves of W that no acyclic non-endelement of W is the sequential limiting set of a sequence of simple closed curves of W. Thus every point of G which is a point and non-endelement of W is a K-curve of G. If h is a non-endelement of G and $W(h)$ is an arc, $W(h)$ is a non-endelement of an interval, H, of arcs of W. It follows from Theorem 3 of [5] that H^* is a closed disc and thus that h is a K-curve of G. In any case, every non-endelement of G is a K-curve. It follows from an argument similar to that in the proof of Lemma 5 that if H is a closed subcollection of arcs of G then for no natural number n is it true that $H^{(n)} = H$. Thus the conditions of Theorem 3 are necessary.

Suppose G satisfies the conditions required of it in the statement of Theorem 3. Since every non-endelement of G is a K-curve then any three simple closed curves of G can be labeled t_1, t_2, and t_3 in such a way that either t_1 is interior to t_2 and t_2 is interior to t_3 or t_1 is interior to t_2 and t_2 and t_3 are mutually exterior. Let a_1 and a_2 be the endelements of G and suppose first that a_2 is a simple closed curve having a_1 in its interior. The continuum a_1 is thus interior to every simple closed curve of $G - a_1$. From the above

remarks and the fact that the non-endelements of G are K-curves it follows that there is a sequence $J_1, J_2, J_3, \cdots$ of mutually exclusive simple closed curves converging to a_1 such that J_1 is a_2 and for each i, a_1 is interior to J_{i+1}, J_{i+1} is interior to J_i and $J_i \cdot G^*$ is an element of G. That the collection W exists now follows by applying Theorem 2 to the intervals of G whose endelements are $J_i \cdot G^*$ and $J_{i+1} \cdot G^*$.

Suppose a_1 is not interior to a_2 and a_2 is not interior to a_1. From the remarks at the beginning of the previous paragraph concerning the simple closed curves of G and the fact that every non-endelement of G is a K-curve it follows that there exist mutually exterior simple closed curves, K_1 and K_2 with exteriors E_1 and E_2 such that a_i is interior to K_i, $(i = 1, 2)$, $K_1 \cdot G^*$ and $K_2 \cdot G^*$ are acyclic elements of G and every element of G between them is acyclic and K_1 and K_2 contain points P_1 and P_2 respectively which are accessible from the same point of $E_1 \cdot E_2 - G^* \cdot E_1 \cdot E_2$. Lemma 6 can be applied to the interval of G whose endelements are $K_1 \cdot G^*$ and $K_2 \cdot G^*$ and the conclusion of the foregoing paragraph can be applied to the remaining two intervals of G. This completes the proof of the sufficiency of the conditions of Theorem 3.

THEOREM 4. *If G is an upper semicontinuous collection of mutually exclusive, compact, atriodic continuous curves which is an arc with respect to its elements then there exists a continuous collection W of mutually exclusive continua filling up a continuous curve in such a way that G is imbedded in W and every element of G is topologically equivalent to the element of W in which it is imbedded if and only if* (1) *each non-endelement of G is a K-curve of G,* (2) *if H is a closed subcollection of arcs of G then for no natural number n is it true that $H^{(n)} = H$,* (3) *the collection of all degenerate elements of G is closed and* (4) *if the simple closed curve or endelement, h, of G is a subset of the boundary of a complementary domain, D, of the sum of the simple closed curves and endelements of G then some point of h is accessible from D.*

Proof. Suppose such a collection, W, exists. The truth of (3) is obvious and it follows from arguments similar to those used in the proof of Theorem 3 that (1) and (2) are true. Since (1) is true, any three simple closed curves of G can be labeled t_1, t_2, t_3 in such a way that either t_1 is interior to t_2 and t_2 is interior to t_3 or t_1 is interior to t_2 and t_2 and t_3 are mutually exterior. Suppose K is the collection of simple closed curves and endelements of G. The collection K is closed. Suppose D is a complementary domain of K^* and let x_1 and x_2 be the elements of K whose sum is the boundary of D.

(If D is bounded by only one element of K, it does not intersect G^* and thus (4) is obvious.) If x_i, $(i = 1, 2)$, is acyclic let $x_i' = x_i$; otherwise let x_i' denote x_i plus its complementary domain which does not interest D. Let U be the collection consisting of x_1', x_2' and all elements of W between x_1 and x_2. The collection U is an arc with respect to its elements, its non-end-elements are acyclic and neither of its endelements separates the plane. Clearly U^* is not the entire plane, it has only one complementary domain, V, and the boundary $B(V)$, of V is a compact continuous curve. It follows from Theorem 3 of [5] that every non-endelement of U intersects $B(V)$ and thus that x_1 and x_2 interesect $B(V)$. Therefore x_1 and x_2 each contain a point accessible from V. But V is a subset of $D - D \cdot G^*$, a connected domain. Therefore x_1 and x_2 contain points accessible from $D - D \cdot G^*$. Thus the condition of Theorem 4 is necessary.

Suppose G is a collection satisfying all the conditions required of it in the statement of Theorem 4. Let L be the collection of degenerate elements of G. If $K + L = G$ it follows from (3) and the fact that every non-endelement of G is a K-curve of G and consequently that K is continuous, (see the proof of Theorem 3 and Theorem II of [1]) that G is continuous and G^* is a continuous curve.

If G contains an arc not in $K + L$ there is a segment of arcs of G whose endelements belong to $K + L$. The collection of all such segments is countable. Denote them by $U_1, U_2, U_3, \cdots$ and denote their respective endelements by x_i and y_i. From condition (4) it follows that there is a sequence $J_1, J_2, J_3, \cdots$ of simple closed curves every convergent subsequence of which converges to a subset of an element of G such that for each i, (1) $J_i \cdot G^*$ is a subset of $x_i + y_i$, $J_i \cdot x_i$ and $J_i \cdot y_i$ are degenerate and $x_i - x_i \cdot J_i$ (if it exists) is not separated in the plane from $y_i - y_i \cdot J_i$ (if it exists) by J_i, (2) U_i^* is a subset of the complementary domain of $J_i + x_i + y_i$ whose boundary is $J_i + x_i + y_i$ and (3) if $p \neq i$ and J_p and J_i intersect, their common part is a point, U_i and U_p have a common endelement and $J_p \cdot J_i$ is a subset of this endelement. Because of the continuity of $K + L$, if the methods of Lemma 6 are now applied to the intervals $x_i y_i$ of G where x_i, y_i, J_i, and D_i replace J_1, J_2, J, and $E_1 \cdot E_2$ respectively of Lemma 6, the proof that the conditions of hTeorem 4 are sufficient is now complete.

GOUCHER COLLEGE.

REFERENCES.

[1] R. D. Anderson, "Continuous collections of continuous curves in the plane," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 647-657.

[2] ———, "On monotone interior mappings in the plane," *Transactions of the American Mathematical Society*, vol. 73 (1952), pp. 211-222.

[3] Herman Cohen, "Some results concerning homogeneous plane continua," *Duke Mathematical Journal*, vol. 18 (1951), pp. 467-474.

[4] Mary-Elizabeth Hamstrom, "Concerning webs in the plane," *Transactions of the American Mathematical Society*, vol. 74 (1953), pp. 500-513.

[5] ———, "Concerning continuous collections of continuous curves," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 240-243.

[6] R. L. Moore, "Concerning webs in the plane," *Proceedings of the National Academy of Sciences*, vol. 29 (1943), pp. 389-393.

[7] ———, "Foundations of Point Set Theory," *American Mathematical Society Colloquium Publications*, vol. XIII, New York, 1932.

ON ALGEBRAIC HOMOGENEOUS SPACES.*

By MORIKUNI GOTO.

1. Introduction. For the sake of simplicity, we call a connected compact complex analytic homogeneous manifold a *C-space,* if the fundamental group is finite, where by homogeneity we mean the existence of a transitive group of biregular automorphisms. The problems on C-spaces, especially the classifications of C-spaces, were studied in detail by H. C. Wang in *Closed manifolds with homogeneous complex structure, American Journal of Mathematics,* vol. 75 (1953), which will be quoted as [Wang] in this paper; Wang defined a C-space as a simply connected C-space in our definition. Here we shall consider a special class of C-spaces.

Let $\mathfrak{R}$ be a connected compact Lie group, and $\mathfrak{U}$ a closed subgroup of $\mathfrak{R}$. It is well-known that all maximal connected abelian subgroups of $\mathfrak{R}$ are conjugate tori, and the Euler characteristic $\chi(\mathfrak{R}/\mathfrak{U})$ of the coset space $\mathfrak{R}/\mathfrak{U}$ does not vanish if and only if $\mathfrak{U}$ contains a maximal torus $\mathfrak{T}$. We shall call a coset space $\mathfrak{R}/\mathfrak{U}$ *semi-simple* if $\chi(\mathfrak{R}/\mathfrak{U}) \neq 0$. Since any C-space is homeomorphic to a coset space of a connected compact Lie group, we obtain a class of semi-simple C-spaces.

Let us denote by Proj. (s) the complex projective group of dimension s, and by Grass. $\binom{r}{n}$ the Grassmann manifold, the manifold of n dimensional linear subspaces in the r dimensional complex vector space. By using Plücker coordinates, Grass. $\binom{r}{n}$ is imbedded naturally in Proj. $\left(\binom{r}{n}-1\right)$ as an algebraic subvariety. A complex analytic manifold V of complex dimension n is called *algebraic* (an algebraic variety) if there exists a biregular mapping from V onto an algebraic subvariety V_1 of a certain dimensional projective space. And if, in particular, we can find V_1 which is birationally equivalent to Proj. (n), we call V *rational* (a rational variety).

The principal result in this paper is the following:

THEOREM. *A C-space is a rational variety if it is semi-simple,* (*and vice versa* [1]).

* Received January 4, 1954.

[1] A. Borel in "Sur les espace homogènes Kähleriens compacts" (to appear) proved that a C-space is semi-simple if it is Kählerian.

Let V be a semi-simple C-space of complex dimension n, homeomorphic to $\mathfrak{R}/\mathfrak{U}$, where $\mathfrak{R}$ is a connected compact Lie group of dimension r. Then, more precisely, *we can give a biregular mapping from V into Grass. $\binom{r}{n}$, in such a way that the image is a subvariety composed of an n-dimensional affine space and a proper subvariety.*

As a special case of the above theorem, the coset space $\mathfrak{R}/\mathfrak{T}$ of a connected compact Lie group modulo a maximal torus has a complex structure of a rational variety.

Now we shall call a semi-simple C-space *simple* if it is homeomorphic to a coset space of a compact simple Lie group. Then, as to the structure of semi-simple C-spaces, we have:

I. *A semi-simple C-space is a direct product of a certain number of simple C-spaces.*

The classification of simple C-spaces was done almost completely in [Wang]. Also as a direct consequence of our theorem and [Wang], we have:

II. *A C-space is a complex analytic fibre bundle with a torus as the fibre over a rational variety.*

The subjects of this paper have connections with those of recent works of A. Borel and A. Weil (to appear). In particular, the fact that a semi-simple C-space is simply connected (§ 3, Proposition 1), was pointed out by Borel. The proof given here is a by-product of our methods and is different from the proof of Borel.

2. Lemmas. 1. Let $\mathfrak{GL}(r)$ be the group of all non-singular matrices of degree r over the field of complex numbers, and let $\mathfrak{N}(r)$ and $\mathfrak{P}(r)$ be subgroups of $\mathfrak{GL}(r)$ composed of all elements of the form

$$\begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ * & & & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} * & & & * \\ & * & & \\ & & \ddots & \\ & & & * \\ 0 & & & * \end{pmatrix}$$

respectively. We call an element $g = (g_{ij})$ in $\mathfrak{GL}(r)$ singular if some of the determinants

$$g_{11}, \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix}, \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix}, \cdots, \begin{vmatrix} g_{11} & g_{12} & \cdots & g_{1\,r-1} \\ g_{21} & g_{22} & \cdots & g_{2\,r-1} \\ \cdots & \cdots & \cdots & \cdots \\ g_{r-1\,1} & g_{r-1\,2} & \cdots & g_{r-1\,r-1} \end{vmatrix}$$

vanishes, and denote by $\mathfrak{S}$ the set of all singular elements. Then it is easy to see that a non-singular element g of $\mathfrak{GL}(r)$ can be decomposed into the product of an element $n(g)$ in $\mathfrak{N}(r)$ and $p(g)$ in $\mathfrak{P}(r)$. In fact, we have:

LEMMA 1. $\mathfrak{GL}(r)-\mathfrak{S}=\mathfrak{N}(r)\mathfrak{P}(r)$,[2] $\mathfrak{N}(r)\cap\mathfrak{P}(r)=1$ (the unit matrix). *The decomposition* $g=n(g)p(g)$ *is unique, and the coefficients of* $n(g)$ *and* $p(g)$ *are rational functions of* g_{ij}.

2. Let G be a complex semi-simple Lie algebra of dimension r. After H. Weyl we can choose a complex basis $h_1,\cdots,h_l, e_\alpha, e_{-\alpha},\cdots, e_\omega, e_{-\omega}$ of G, so that $[h_i, h_j]=0$ $(i,j=1,\cdots,l)$,

$$[\sum_{i=1}\lambda_i h_i, e_\alpha]=(\alpha\lambda)e_\alpha, \qquad (\alpha\lambda)=\sum_{i=1}^{l}\alpha_i\lambda_i$$

$$[e_\alpha, e_{-\alpha}]=\sum_{i=1}^{l}\alpha_i h_i, \qquad [e_\alpha, e_\beta]=N_{\alpha\beta}e_{\alpha+\beta}.$$

Here $\alpha=(\alpha_1,\cdots,\alpha_l),\cdots$ are real vectors, called root vectors, and each $N_{\alpha\beta}$ is a real number, which does not vanish if $\alpha+\beta$ is a root vector. If $\alpha+\beta$ is not a root vector, we may put $N_{\alpha\beta}=e_{\alpha+\beta}=0$.

We define a linear order in the l dimensional real vector space R^l, so that R^l becomes an ordered abelian group. We put namely

$$(\lambda_1,\cdots,\lambda_l)>(\lambda'_1,\cdots,\lambda'_l)$$

if $\lambda_1=\lambda'_1,\cdots,\lambda_{i-1}=\lambda'_{i-1},\lambda_i>\lambda'_i$ for some i. Then we can decompose G into a direct module sum of two subalgebras defined by

$$N=Ce_{-\alpha}+\cdots+Ce_{-\omega}, \qquad P=Ch_1+\cdots+Ch_l+Ce_\alpha+\cdots+Ce_\omega,$$

$(\alpha>0,\cdots,\omega>0)$ where C denotes the field of complex numbers.

Let x be an element of G. We denote by $\mathrm{ad}x$ the matrix representing the mapping $\mathrm{ad}x\cdot y=[x,y]$, $y\in G$, with respect to the basis

$$e_\alpha,\cdots,e_\omega,h_1,\cdots,h_l,e_{-\omega},\cdots,e_{-\alpha},$$

$$(\alpha>\cdots>\omega>0>-\omega>\cdots>-\alpha).$$

The correspondence $x\to\mathrm{ad}x$ gives a faithful representation of G. We use a Latin capital * for a complex subalgebra of G, and use the notation $\mathrm{ad}^*=\{\mathrm{ad}x\,|\,x\in *\}$, and the corresponding German capital denotes the matric Lie group generated by ad*. For example, $\mathfrak{G}$ is the Lie group, generated by $\mathrm{ad}G$, namely the adjoint group of G.

[2] This decomposition was used by I. Gelfand in his theory of unitary representations of Lie groups.

LEMMA 2. $\mathfrak{G} \cap \mathfrak{N}(r) = \mathfrak{N} = \{n(g) \mid g \in \mathfrak{G} - \mathfrak{S}\}$

$\mathfrak{G} \cap \mathfrak{P}(r) = \mathfrak{P} = \{p(g) \mid g \in \mathfrak{G} - \mathfrak{S}\}$

$\mathfrak{G} - \mathfrak{S} = \mathfrak{N}\mathfrak{P}, \qquad \mathfrak{N} \cap \mathfrak{P} = 1.$

Proof. Since $G = N + P$, $N \cap P = 0$, we have $\mathrm{ad}G = \mathrm{ad}N + \mathrm{ad}P$, $\mathrm{ad}N \cap \mathrm{ad}P = 0$. Therefore we can take nuclei N_l and P_l of the Lie groups $\mathfrak{N}$ and $\mathfrak{P}$, integrated by $\mathrm{ad}N$ and $\mathrm{ad}P$ respectively, so that $\mathfrak{G}_l = \mathfrak{N}_l\mathfrak{P}_l$ $(\mathfrak{N}_l \cap \mathfrak{P}_l = 1)$ is a nucleus of $\mathfrak{G}$, where by a nucleus we mean a neighborhood of the identity. Here, by our choice of the basis the matrices in $\mathrm{ad}N$ and $\mathrm{ad}P$ are of the forms

$$\begin{pmatrix} 0 & & & 0 \\ & 0 & & \\ & & \ddots & \\ * & & & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} * & & & * \\ & * & & \\ & & \ddots & \\ 0 & & & * \end{pmatrix}$$

respectively, whence we have $\mathfrak{N} \subseteq \mathfrak{N}(r)$, $\mathfrak{P} \subseteq \mathfrak{P}(r)$.

Now it is known that $\mathfrak{G}$ coincides with the identity component of the group of all automorphisms of G, and therefore algebraic. In fact we can find polynomials $\phi_1, \phi_2, \cdots$ with respect to r^2 variables ξ_{ij} $(i, j = 1, \cdots, r)$, so that

$$\phi_1(\xi_{ij}) = \phi_2(\xi_{ij}) = \cdots = 0 \quad \text{imply} \quad (\xi_{ij}) \in G.$$

For g in $\mathfrak{G}_l$, $n(g)$ is contained in $\mathfrak{N}$. Hence

$$\phi_1(n(g)) = \phi_2(n(g)) = \cdots = 0, \qquad g \in \mathfrak{G}_l.$$

On the other hand, since $\mathfrak{G} \cap \mathfrak{S}$ is a proper subvariety of $\mathfrak{G}$, the complex dimension of $\mathfrak{G} \cap \mathfrak{S}$ is less than r, whence $\mathfrak{G} - \mathfrak{S}$ is a connected manifold. Therefore the equalities must be valid everywhere in $\mathfrak{G} - \mathfrak{S}$. Hence $n(g) \in \mathfrak{G}$, namely $\{n(g) \mid g \in \mathfrak{G} - \mathfrak{S}\} \subseteq \mathfrak{G} \cap \mathfrak{N}(r)$. The converse implication is obvious. Thus we get $\mathfrak{G} \cap \mathfrak{N}(r) = \{n(g) \mid g \in \mathfrak{G} - \mathfrak{S}\}$.

The left hand side is a closed subgroup, which is a connected set by the equality. Differentiating the closed connected subgroup $\mathfrak{N}(r) \cap \mathfrak{G}$, we get clearly the Lie algebra $\mathrm{ad}N$.

By the same argument, we have the relations on $\mathfrak{P}$.

3. Let Δ be the set of root vectors of G, and M a linear subspace of R^l, and let Φ be the set of root vectors in M: $\Phi = \Delta \cap M$. Denote by $\Phi^\perp$ the complementary set of Φ in Δ: $\Delta - \Phi = \Phi^\perp$. Then, Φ contains $-\beta$ with β, and if β and β' are in Φ and if $\beta + \beta'$ is a root vector, then $\beta + \beta' \in \Phi$.

Moreover, that $\alpha \varepsilon \Phi^{\perp}$ and $\beta \varepsilon \Phi$ implies $\alpha + \beta \varepsilon \Phi^{\perp}$, if it is a root vector.

We may assume that

$$\alpha(1) > \cdots > \alpha(n) > \beta(1) > \cdots > \beta(m) > 0 > -\beta(m) > \cdots > -\beta(1) > -\alpha(n) > \cdots > -\alpha(1),$$

where $\pm \alpha(i) \varepsilon \Phi^{\perp}$, $\pm \beta(j) \varepsilon \Phi$. In fact any basis of R^l such that the last part of which constitutes a basis of M gives a desired ordering of R^l.

Consider the following subalgebras of G:

$$P_1 = Ce_{\alpha(1)} + \cdots + Ce_{\alpha(n)},$$

$$P_2 = Ce_{\beta(1)} + \cdots + Ce_{\beta(m)} + Ch_1 + \cdots + Ch_l$$

$$N_1 = Ce_{-\alpha(1)} + \cdots + Ce_{-\alpha(n)}, \qquad N_2 = Ce_{-\beta(1)} + \cdots + Ce_{-\beta(m)}$$

$$P = P_1 + P_2, \qquad N = N_1 + N_2, \qquad H = P + N_2.$$

Then we have $G = N_1 + H$, $N_1 \cap H = 0$, and $[H, H] = P_1 + [P_2 + N_2, P_2 + N_2]$, where $[P_2 + N_2, P_2 + N_2]$ is a semi-simple subalgebra corresponding to the root system Φ, and P_1 the radical (maximal solvable ideal) of $[H, H]$.

Since P_1 is an ideal of H, we have $\operatorname{ad} H \cdot P_1 \subseteq P_1$. Conversely, let x be an element of G so that $\operatorname{ad} x \cdot P_1 \subseteq P$. We may put $x = y + \sum_{j=1}^{n} \xi_j e_{-\alpha(j)}$, $y \varepsilon H$. Because $[e_{-\alpha(i)}, e_{\alpha(i)}]$ is a non-vanishing linear combination of h's and $[\sum_{j=1}^{n} \xi_j e_{-\alpha(j)}, e_{\alpha(j)}] = \xi_i [e_{-\alpha(i)}, e_{\alpha(i)}] +$ (a linear combination of e's), ξ_i must be zero, namely $x \varepsilon H$. Thus we have $H = \{x \mid \operatorname{ad} x \cdot P_1 \subseteq P_1\}$.

Let $\mathfrak{H}(r)$ be the subgroup of $\mathfrak{GL}(r)$ defined by $\mathfrak{H}(r) = \{g \mid g \cdot P_1 = P_1\}$. Then the Lie algebra of $\mathfrak{H}(r) \cap \mathfrak{G}$ must be $\operatorname{ad} H$. Therefore the Lie subgroup $\mathfrak{H}$ corresponding to $\operatorname{ad} H$ is the identity component of $\mathfrak{H}(r) \cap \mathfrak{G}$. Now let $n(\mathfrak{H})$ be the normalizer of $\mathfrak{H}$ in $\mathfrak{G}$, and g an element of $n(\mathfrak{H})$. Since P_1 is the radical of $[H, H]$, $\mathfrak{P}_1$ is a characteristic subgroup of $\mathfrak{H}$. Therefore $g^{-1}\mathfrak{P}_1 g = \mathfrak{P}_1$, namely $g \cdot P_1 = P_1$, whence $g \varepsilon \mathfrak{H}(r)$. Since $\mathfrak{H}$ is normal in $\mathfrak{H}(r) \cap \mathfrak{G}$, we have $n(\mathfrak{H}) = \mathfrak{H}(r) \cap \mathfrak{G}$.

4. Since $N = N_1 + N_2$ and N_1 is an ideal of N, we have $\mathfrak{N} = \mathfrak{N}_1 \mathfrak{N}_2$, where $\mathfrak{N}_i$ $(i = 1, 2)$ is a Lie subgroup corresponding to $\operatorname{ad} N_i$. On the other hand, by Lemma 2 we know that $\mathfrak{G} - \mathfrak{S} = \mathfrak{N}\mathfrak{P}$. Therefore we have $\mathfrak{G} - \mathfrak{S} = (\mathfrak{N}_1 \mathfrak{N}_2)\mathfrak{P} = \mathfrak{N}_1 (\mathfrak{N}_2 \mathfrak{P}) \subseteq \mathfrak{N}_1 \mathfrak{H}$.

Now, let $N(r)$ be the matric Lie algebra of $\mathfrak{N}(r)$. For x in $N(r)$, $\exp x = 1 + x + x^2/2! + \cdots$ is a polynomial of x of degree less than r.

Conversely let g be in $\mathfrak{N}(r)$. Then $\log g = (g-1) - (g-1)^2/2 + \cdots$ is also a polynomial of g degree less than r, and we have $\exp(\log g) = g$, $\log(\exp x) = x$. Thus the mapping

$$N(r) \ni x = (x_{ij}) \to \exp x = 1 + (y_{ij}) \in \mathfrak{N}(r), \qquad i > j$$

is one-to-one and onto, where each y_{ij} is given as a polynomial of x's and vice versa.

Let g be an element in $\mathfrak{N}_1 \cap n(\mathfrak{H})$. Since g is contained in $\mathfrak{N}_1$, we have $g = \exp(\operatorname{ad} x)$, $x \in N_1$, where $\operatorname{ad} x$ is a polynomial of g. Since $g \cdot P_1 = P_1$, we have $\operatorname{ad} x \cdot P_1 \subseteq P_1$, whence $x \in H$. Since $N_1 \cap H = 0$, we have $x = 0$, namely $g = 1$. Thus we have $\mathfrak{N}_1 \cap n(\mathfrak{H}) = 1$.

Suppose that $n(\mathfrak{H})$ is not connected, and let us decompose $n(\mathfrak{H})$ into cosets modulo $\mathfrak{H}$: $n(\mathfrak{H}) = \mathfrak{H} \dotplus \mathfrak{H}g_1 \dotplus \cdots$, which yields $\mathfrak{N}_1 n(\mathfrak{H}) = \mathfrak{N}_1\mathfrak{H} \dotplus \mathfrak{N}_1\mathfrak{H}g_1 \dotplus \cdots$, where $\dotplus$ denotes the direct sum of sets. Since $\mathfrak{N}_1\mathfrak{H}$ contains an open set, so does $\mathfrak{N}_1\mathfrak{H}g_1$. On the other hand

$$\mathfrak{N}_1\mathfrak{H}g_1 \subseteq \mathfrak{S}' = \mathfrak{G} - \mathfrak{N}_1\mathfrak{H} \subseteq \mathfrak{G} \cap \mathfrak{S},$$

where $\mathfrak{G} \cap \mathfrak{S}$ has a lower complex dimension than $\mathfrak{G}$. This is a contradiction. Therefore $\mathfrak{H}$ must coincide with $n(\mathfrak{H})$.

LEMMA 3. $\mathfrak{H} = n(\mathfrak{H}) = \mathfrak{G} \cap \mathfrak{H}(r)$. $\mathfrak{G} - \mathfrak{S}' = \mathfrak{N}_1\mathfrak{H}$, $\mathfrak{N}_1 \cap \mathfrak{H} = 1$, *where $\mathfrak{S}'$ is contained in a closed set of lower complex dimension.*

5. LEMMA 4. *The complex coset space $\mathfrak{GL}(r)/\mathfrak{H}(r)$ is biregularly equivalent to Grass.* $\binom{r}{n}$.

Proof. Consider the mapping f defined by

$$\mathfrak{GL}(r) \ni g = (g_1 g_2 \cdots g_r) \to [g_1 \wedge g_2 \wedge \cdots \wedge g_n] \in \text{Grass.}\binom{r}{n},$$

where

$$g_1 = \begin{pmatrix} g_{11} \\ \vdots \\ g_{r1} \end{pmatrix}, \cdots, \quad g_r = \begin{pmatrix} g_{1r} \\ \vdots \\ g_{rr} \end{pmatrix},$$

and $[g_1 \wedge \cdots \wedge g_n]$ denotes the element of Grass. $\binom{r}{n}$, spanned by $g_1, \cdots, g_n$.

This mapping is onto, and two elements g and $g' = (g'_1 g'_2 \cdots g'_r)$ go to the same point if and only if

$$g'_i = a_{i1}g_1 + \cdots + a_{in}g_n \qquad (i = 1, 2, \cdots, n),$$

namely $g' = gh$ $h \in \mathfrak{H}(r)$. Hence f is one-to-one on the left coset space $\mathfrak{GL}(r)/\mathfrak{H}(r)$, and is clearly biregular.

3. Propositions. We shall start from a C-space V of complex dimension n, and first mention the results in [Wang] in a way which is convenient for us:

There exists a connected complex semi-simple Lie group $\tilde{\mathfrak{G}}$ of biregular automorphisms transitive on V. Therefore V is biregularly equivalent to a coset space $\tilde{\mathfrak{G}}/\tilde{\mathfrak{L}}$, where $\tilde{\mathfrak{L}}$ is a closed complex subgroup of $\tilde{\mathfrak{G}}$. Let G be the Lie algebra of $\tilde{\mathfrak{G}}$, and L that of $\tilde{\mathfrak{L}}$. Then we can find a subalgebra H which satisfies the conditions in § 2. 3., so that $G \supset H \supset L$ and $[H, H] = [L, L]$. Here H coincides with L if and only if V is semi-simple.

Let $\mathfrak{D}$ be the (finite) center of $\tilde{\mathfrak{G}}$. Then the factor group $\tilde{\mathfrak{G}}/\mathfrak{D}$ is isomorphic to the adjoint group $\mathfrak{G}$. Let $\mathfrak{H}$ be the closed Lie subgroup corresponding to H in $\mathfrak{G}$. Since $n(\mathfrak{H}) = \mathfrak{H}$ by Lemma 3, and since $\mathfrak{D}$ is known to be contained in $\tilde{\mathfrak{H}}$, the normalizer $n(\tilde{\mathfrak{H}})$ of $\tilde{\mathfrak{H}}$ coincides with $\tilde{\mathfrak{H}}$, and $\tilde{\mathfrak{G}}/\tilde{\mathfrak{H}}$ is biregularly equivalent to $\mathfrak{G}/\mathfrak{H}$.

Let V be semi-simple. Then since $\tilde{\mathfrak{L}}$ coincides with $\tilde{\mathfrak{H}}$ locally, $n(\tilde{\mathfrak{H}}) = \tilde{\mathfrak{H}}$ implies $\tilde{\mathfrak{L}} = \tilde{\mathfrak{H}}$. On the other hand $\tilde{\mathfrak{G}}/\tilde{\mathfrak{H}}$ is known to be simply connected. Thus we have:

PROPOSITION 1. *A semi-simple C-space is simply connected.* (See the end of § 1.)

Now we shall prove the following:

PROPOSITION 2. *A semi-simple C-space V is a rational variety, containing a complex affine space V_0 such that $V - V_0$ is a subvariety. Moreover there exists a biregular mapping from V into a projective space which is rational in terms of a system of affine coordinates on V_0.*

Proof. We may assume $V = \mathfrak{G}/\mathfrak{H}$. By Lemma 4 we have a biregular mapping f from $\mathfrak{G}\mathfrak{Q}(r)/\mathfrak{H}(r)$ onto Grass. $\binom{r}{n}$. Since $\mathfrak{G} \cap \mathfrak{H}(r) = \mathfrak{H}$ by Lemma 3, f is one-to-one on $\mathfrak{G}/\mathfrak{H}$. Therefore f induces a one-to-one biregular imbedding of V into Grass. $\binom{r}{n}$.

Now by the decomposition $\mathfrak{G} - \mathfrak{S}' = \mathfrak{N}_1\mathfrak{H}$, V contains an open submanifold $V_0 = f(N_1)$. Let $\xi_1, \cdots, \xi_n$ be the parameters of the Lie algebra N_1. Then the Plücker coordinates of $f(V_0)$ are polynomials of $\xi_1, \cdots, \xi_n$, say $f_0(\xi), \cdots, f_N(\xi)$, where $N = \binom{r}{n} - 1$. Let us construct the irreducible variety with the generic point $(f_i(\xi))$. It is clear that the variety coincides with $f(V)$.

Thus we obtained a rational mapping from Proj. (n) onto $f(V)$. Since the algebraic correspondence given by $\xi \to (f_i(\xi))$ is one-to-one for generic

points ξ, the mapping is birational, whence $f(V)$ is a rational variety. The image of the ∞-hyperplane in Proj. (n) coincides with $f(V) - f(V_0)$.

PROPOSITION 3. *A C-space is a complex analytic fibre bundle over a rational variety with a torus as the fibre.*

Proof. We retain the above notations. Let $V = \tilde{\mathfrak{G}}/\tilde{\mathfrak{L}}$ be a C-space, and g an element of $\tilde{\mathfrak{L}}$. Since P_1 (see § 2. 3.) is the radical of the commutator subalgebra $[H, H] = [L, L]$, the corresponding Lie subgroup is a characteristic subgroup of the identity component of L. Therefore $g^{-1}\mathfrak{P}_1 g = \mathfrak{P}_1$, namely $gP_1 = P_1$, whence g is contained in $\tilde{\mathfrak{H}}$. Thus, $\tilde{\mathfrak{L}}$ is a subgroup of $\tilde{\mathfrak{H}}$. Since $\tilde{\mathfrak{L}}$ contains the commutator subgroup of $\tilde{\mathfrak{H}}$, $\tilde{\mathfrak{L}}$ is a normal subgroup of $\tilde{\mathfrak{H}}$, and the factor group $\tilde{\mathfrak{H}}/\tilde{\mathfrak{L}}$ is abelian. Hence $\mathfrak{H}/\mathfrak{L}$ is considered as a complex toral group operating on $V = \tilde{\mathfrak{G}}/\tilde{\mathfrak{L}}$, and we have: $V/\tilde{\mathfrak{H}}/\tilde{\mathfrak{L}} = \tilde{\mathfrak{G}}/\tilde{\mathfrak{H}}$, which establishes the proposition.

PROPOSITION 4. *A semi-simple C-space is a direct product of simple C-spaces.*

Proof. Let $\mathfrak{G}/\mathfrak{H}$ be a semi-simple C-space, and let $G = G_1 + \cdots + G_s$ be the decomposition of the Lie algebra G into simple ideals. Then clearly $H = H_1 + \cdots + H_s$, where H_i is a subalgebra of G_i. $(i = 1, 2, \cdots, s.)$ Let $\mathfrak{G}_i$ be the adjoint group of G_i, and $\mathfrak{H}_i$ the Lie subgroup of G_i corresponding to $\mathrm{ad}\, H_i$. By Proposition 1, $\mathfrak{H}$ is connected. Since $\mathfrak{G} = \mathfrak{G}_1 \times \cdots \times \mathfrak{G}_s$, we have $\mathfrak{H} = \mathfrak{H}_1 \times \cdots \times \mathfrak{H}_s$, whence $\mathfrak{G}/\mathfrak{H} = \mathfrak{G}_1/\mathfrak{H}_1 \times \cdots \times \mathfrak{G}_s/\mathfrak{H}_s$.

These propositions clearly imply the statements mentioned in § 1.

THE INSTITUTE FOR ADVANCED STUDY,
AND TOKYO UNIVERSITY OF EDUCATION.

NUMBER OF POINTS OF VARIETIES IN FINITE FIELDS.*

By SERGE LANG and ANDRÉ WEIL.

1. Statement and proof of the main theorem. Let V be a variety in a projective space P^n. If V has dimension r and degree d, we shall indicate this by writing $V = V_{n,d,r}$. A point of V can be represented by homogeneous coordinates $(x) = (x_0, \cdots, x_n)$, and we shall also say that (x) is a point of V. We use k to denote a finite field and q to denote the number of elements in k. Let V be defined over k. A point (x) of V is said to be rational over k, or more briefly to be in k, if its coordinate ratios x_i/x_j $(x_j \neq 0)$ are in k. We intend to give an estimate for the number of points of V which are in k. Denoting this number by N, we prove

THEOREM 1. *There exists a constant $A(n, d, r)$ depending only on n, d, r such that for any variety $V = V_{n,d,r}$ defined over a finite field k we have*

$$| N - q^r | \leqq \delta q^{r-\frac{1}{2}} + A(n, d, r) q^{r-1}$$

where $\delta = (d-1)(d-2)$.

If $r = 1$, i.e. if V is a curve, the above diophantine statement is a reformulation of the Riemann hypothesis in function fields [2]. Indeed, let V_1 be a non singular projective curve, defined over k, birationally equivalent to V over k. Let g be the genus of V and V_1. Then we know that the number N_1 of rational points of V_1 satisfies $| 1 + q - N_1 | \leqq 2gq^{\frac{1}{2}}$. Each non singular point of V corresponds to exactly one point of V_1. The number of singular points on V is bounded by a constant depending only on d, and each singular point of V corresponds at most to d points on V_1. Hence $| N_1 - N | \leqq A$ where A is a constant depending only on d. This shows that $| N - q | \leqq 2gq^{\frac{1}{2}} + (A + 1)$. Using the fact that $g \leqq \frac{1}{2}(d-1)(d-2)$, we see that Theorem 1 is true for $r = 1$.

The proof of Theorem 1 will now be carried out by induction on the dimension r, and the arguments will be of an elementary nature.

We begin by two lemmas and use throughout the terminology of [1],

* Received November 30, 1953.

except that a bunch of varieties will be called an algebraic set. We shall say that an algebraic set is defined over k if it is defined by algebraic equations $f_\alpha = 0$ with coefficients in k. If the algebraic set is in projective space, then the f_α can be chosen to be forms, and the algebraic set is identified with the rays of zeros of these forms. For example, a hyperplane H in P^n is defined by an equation $\sum w_i X_i = 0$. If $w_i \in k$, then H is defined over k, or as we shall say more briefly, H is in k.

Let Z be a positive cycle in P^n, of degree $d > 0$, and dimension r. We can then write $Z = \sum a_i V_i$ as a formal sum of distinct varieties V_i, with integer coefficients $a_i > 0$. We then have $d = \deg Z = \sum a_i \cdot \deg V_i$, and $\dim V_i = r$ for each i. By a point of Z we shall mean a point of any of the varieties V_i. The set of points of Z is an algebraic set denoted by $|Z|$. If Z is rational over k, then $|Z|$ is defined over k. We let N_Z be the number of points of Z in k.

LEMMA 1. *There exists a constant $A_1(n, d, r)$ depending only on n, d, r such that for any positive cycle Z in P^n, of degree d, dimension r, and rational over k, we have $N_Z \leqq A_1 q^r$.*

Proof. By induction on the dimension r. Suppose first $r = 0$. Then $|Z|$ consists of at most d points, and the lemma is trivial. Assume now $r \geqq 1$. If we express Z as a sum of prime rational cycles over k, then there will be at most d such cycles in the sum. Hence it suffices to prove the result for a prime rational cycle. Let $(X_0, \cdots, X_n)$ be the variables of P^n. For some pair of indices, say $(0, 1)$, Z (which we assume to be prime rational over k) intersects properly the hyperplanes H_ξ defined by the equations $X_0 - \xi X_1 = 0$ where ξ ranges over k, and the hyperplane H defined by $X_1 = 0$. Every point of Z in k is contained in one of the cycles $Z \cdot H_\xi$ or $Z \cdot H$, each of which has degree d by Bezout's Theorem ([1], App. I) and is of dimension $r - 1$. According to the induction hypothesis, there exists a constant $B(n, d, r-1)$ such that $N_{Z \cdot H_\xi}$ and $N_{Z \cdot H} \leqq B(n, d, r-1) q^{r-1}$. As we have exactly $q + 1$ cycles, we see that there are at most $B(q+1) q^{r-1}$ points in Z, and this is certainly $\leqq 2Bq^r$. The constant $2B$ is what we are looking for.

Our second lemma will be concerned with the following situation. Let V be a variety in P^n, of dimension $r \geqq 2$, and not contained in a hyperplane. Let P'^n be the projective space dual to P^n, and let $(w) = (w_0, \cdots, w_n)$ be the homogeneous coordinates of a point in P'^n. Let H_w be the hyperplane defined by $\sum w_i X_i = 0$. Then the set R of points (w) of P' such that the cycle $V \cdot H_w$ is not a variety is an algebraic set, defined over every field of definition of V, and $R \neq P'$. (Cf. [4], § 1, 6). The hyperplanes H such

that $V \cdot H$ is not a variety may therefore be viewed as forming an algebraic set R in P'. Lemma 2 gives an estimate for the number of such hyperplanes defined over a finite field k.

LEMMA 2. *There exists a constant $A_2(n, d, r)$ depending only on n, d, r having the following property. If $V = V_{n,d,r}$ is any variety defined over k, not contained in a hyperplane, and if N_R is the number of hyperplanes H in k such that $V \cdot H$ is not a variety, then $N_R \leqq A_2 q^{n-1}$.*

Proof. Let Z be a positive cycle of dimension r and degree d in P^n. Let $F(U^{(0)}, \cdots, U^{(r)})$ be the associated form of Z. It is of degree d in each set of variables $U^{(0)}, \cdots, U^{(r)}$, and its coefficients are called the Chow coordinates of Z. These are viewed as the homogeneous coordinates of a point in a projective space P^M. The coordinates $(c) = (c_0, \cdots, c_M)$ of a cycle which is not a variety form an algebraic set C in P^M (cf. [4], § 1, 6), and C is defined over the prime field. We let $\phi_\alpha = 0$ be a finite system of equations with coefficients in the prime field, defining C. The forms ϕ_α depend only on n, r, and d.

Let now $V = V_{n,d,r}$ be a variety defined over k, not contained in a hyperplane. Let (w) be a point of P', and let $c(w) = (c_0(w), \cdots, c_M(w))$ be the Chow coordinates of the cycle $V \cdot H_w$. If $F(U^{(0)}, \cdots, U^{(r)})$ is the associated form of V, then it can be verified that the associated form of $V \cdot H_w$ is $F(w, U^{(0)}, \cdots, U^{(r-1)})$. Hence each $c_i(w)$ is a form of degree d in the quantities (w), with coefficients in k.

Let R be the algebraic set of points (w) in P' such that $V \cdot H_w$ is not a variety. This algebraic set R is defined over k, and we have $(w) \in R$ if and only if $c(w) \in C$, i. e. if and only if $\phi_\alpha(c(w)) = \phi_\alpha(c_0(w), \cdots, c_M(w)) = 0$ for all α. If $(W) = (W_0, \cdots, W_n)$ are the variables of P', at least one of the polynomials $\phi_\alpha(c(W))$ does not vanish identically. (If they all did, R would be the entire space P', which is not true.) The degree of this $\phi_\alpha(c(W))$ depends only on the degree of ϕ_α and on the degree d of each $c_i(W)$. Since the ϕ_α depend only on n, d, and r we have proved that R is contained in a hypersurface $\phi_\alpha(c(W)) = 0$ with coefficients in k, whose degree is a constant e depending only on n, d, r. This hypersurface defines a positive cycle Z in P^n, of degree e and dimension $n - 1$, rational over k. By Lemma 1, we have $N_R \leqq N_Z \leqq A_1(n, e, n-1) q^{n-1}$ where A_1 is the constant of Lemma 1. Putting $A_2(n, d, r) = A_1(n, e, n-1)$, we see that Lemma 2 is proved.

Remark. All the constants depending on n, d, r which we are finding in this paper can easily be estimated, except for e in the proof of the preceding lemma.

We are now in a position to prove the main theorem.

Let $V = V_{n,d,r}$ be defined over k. We may assume that V is not contained in a hyperplane. Indeed, if V is contained in a hyperplane, or in a linear variety, then it is easily verified that the smallest linear variety L_0 containing V is also defined over k. (The linear forms defining L_0 are precisely the linear forms contained in the prime homogeneous ideal in $k[X]$ defining V.) L_0 can then be viewed as a projective space of dimension $n_0 \leqq n$, and it would suffice to prove our theorem for V as a subvariety of L_0.

If V is not contained in a hyperplane, then the cycle $V \cdot H$ is defined for every hyperplane H, and Lemma 2 will be applicable.

We consider the pairs $((x), H)$ consisting of a point (x) of V in k and a hyperplane H in P'^n defined by an equation $\sum w_i X_i = 0$ with $w_i \in k$ and passing through (x), i. e. such that $H(x) = \sum w_i x_i = 0$. We shall count these pairs in two ways.

We shall denote by κ_{n+1} the number of points in k in the projective space P^n. This is given by $\kappa_{n+1} = q^{n+1} - 1/q - 1 = 1 + q + \cdots + q^n$. This is also the number of hyperplanes H of P' in k. Similarly, the number of hyperplanes in k passing through a given point (x) in k is κ_n. From these remarks, it follows that the number of pairs $((x), H)$ is given by the following two equal numbers:

$$N\kappa_n = \sum_H N_{V \cdot H}$$

where N is the number of points of V in k, $N_{V \cdot H}$ is the number of points of the cycle $V \cdot H$ in k, and the sum is taken over all hyperplanes H in k. Solving for N, and using the same notation as in Lemma 2, we see that N is equal to

$$(1) \qquad \frac{1}{\kappa_n} \sum_{H \notin R} N_{V \cdot H} + \frac{1}{\kappa_n} \sum_{H \in R} N_{V \cdot H}$$

where R is the algebraic set of hyperplanes H such that $V \cdot H$ is not a variety.

If N_R is the number of terms in the second sum, i. e. the number of hyperplanes H in R, the number of terms in the first sum is $\kappa_{n+1} - N_R$. We shall now estimate the two sums, and begin with the first one.

For $H \notin R$, $V \cdot H$ is a variety of degree d, dimension $r - 1$, defined over k. It follows from the induction hypothesis that for $H \notin R$, we have $| N_{V \cdot H} - q^{r-1} | \leqq \delta q^{r-3/2} + A_3 q^{r-2}$ where $A_3 = A(n, d, r-1)$ is the constant determined inductively, depending only on n, d, and $r - 1$. Summing over $H \notin R$, and dividing by κ_n we get

$$\left| \frac{1}{\kappa_n} \sum_{H \notin R} N_{V \cdot H} - Q \right| \leqq \frac{\kappa_{n+1} - N_R}{\kappa_n} (\delta q^{r-3/2} + A_3 q^{r-2})$$

where $Q = \kappa_n^{-1}(\kappa_n - N_R)q^{r-1}$. We note that $\kappa_{n+1} = q\kappa_n + 1$, and $q^{n-1} \leqq \kappa_n$. Also we know by Lemma 2 that $N_R \leqq A_2 q^{n-1} \leqq A_2\kappa_n$. This shows that $|Q - q^r| \leqq (1 + A_2)q^{r-1}$ and that $\kappa_n^{-1}(\kappa_{n+1} - N_R) \leqq \kappa_n^{-1}\kappa_{n+1} \leqq q + 1$. From this we conclude immediately that

$$|\frac{1}{\kappa_n}\sum_{H \notin R} N_{V\cdot H} - q^r| \leqq \delta q^{r-\frac{1}{2}} + A_4 q^{r-1} \tag{2}$$

with a suitable A_4.

We now turn to our second sum, $\sum_{H \in R} N_{V\cdot H}$, and we shall prove that it can be absorbed by the error term involving q^{r-1} only.

If $H \in R$, then $V \cdot H$ is a cycle of degree d, dimension $r - 1$, and rational over k. By Lemma 1, it follows that $N_{V\cdot H} \leqq A_1 q^{r-1}$ for a suitable constant A_1. By Lemma 2, we know that $N_R \leqq A_2 q^{n-1}$. Hence

$$\frac{1}{\kappa_n}\sum_{H \in R} N_{V\cdot H} \leqq \frac{q^{n-1}}{\kappa_n} A_1 A_2 q^{r-1} \leqq A_5 q^{r-1} \tag{3}$$

because $q^{n-1} \leqq \kappa_n$, as we have already remarked.

Combining (1), (2), and (3) we see that there exists a constant A_6 (the desired constant $A(n, d, r)$) such that $|N - q^r| \leqq \delta q^{r-\frac{1}{2}} + A_6 q^{r-1}$. This concludes the proof of the theorem.

2. Corollaries and applications. We list some of the immediate corollaries to Theorem 1.

In the first place, our theorem is of the nature of an asymptotic result. If k_ν is the extension of degree ν over k, and $N^{(\nu)}$ is the number of points of V in k_ν then

$$N^{(\nu)} = q^{\nu r} + O(q^{\nu(r-\frac{1}{2})}) \quad \text{for} \quad \nu \to \infty.$$

In particular, if $\nu \to \infty$ then $N^{(\nu)} \to \infty$ also. We shall now see that this asymptotic behavior can be stated more generally for abstract varieties.

Let $V = V_{n,d,r}$ be a variety defined over k, and let F be a frontier on V (i. e. an algebraic set properly contained in V) also defined over k. Let $N^{(\nu)}{}_{V-F} = N^{(\nu)}{}_V - N^{(\nu)}{}_F$ be the number of points of $V - F$ which are in k_ν. Then from Lemma 1 we see immediately that $|N^{(\nu)}{}_{V-F} - N^{(\nu)}{}_V| \leqq Bq^{\nu(r-1)}$ where B is a constant depending on V and on F.

If V and V' are two varieties of dimension r in P^n, defined over k, and if T is a birational correspondence between them, also defined over k, then there exist frontiers F and F' defined over k such that T is everywhere biregular in $V - F$ and $V' - F'$. T gives a $1 - 1$ correspondence between the points of $V - F$ and those of $V' - F'$, so $N^{(\nu)}{}_{V-F} = N^{(\nu)}{}_{V'-F'}$.

Let now $V = (V_\alpha, F_\alpha, T_{\beta\alpha})$ be an abstract variety, represented by varieties V_α in projective space, with frontiers F_α, and birational correspondences $T_{\beta\alpha}$, all of which are defined over k. If P is a point of V, then P has representatives P_α in some of the V_α. If one of these P_α is rational over k, then so are all the other representatives, and hence in this case we may say that P is rational over k.

If K is a function field of dimension r over k, then we shall say that an abstract variety V is a model of K if V is defined over k, and if $K = k(P)$ where P is a generic point of V over k. The following two corollaries of Theorem 1 are immediate consequences of the preceding discussion.

COROLLARY 1. *Let K be a function field of dimension r over a finite constant field k. If V is any abstract model of K, and $N^{(\nu)}$ the number of points of V in k_ν, then $N^{(\nu)} - q^{\nu r}$, mod $O(q^{\nu(r-1)})$, is a birational invariant (i. e. does not depend on the choice of the model V of K).*

COROLLARY 2. *Let K/k be as in Corollary 1. There exists a constant γ such that for any abstract model V of K, we have*

$$| N^{(\nu)} - q^{\nu r} | \leqq \gamma q^{\nu(r-\frac{1}{2})} + B q^{\nu(r-1)}$$

where B is a constant depending on V. If K has a projective model of degree d, then we can take $\gamma \leqq (d-1)(d-2)$.

The smallest constant γ that can be selected in Corollary 2 is obviously a birational invariant, and we shall return to discuss it below. For the moment, we note that an abstract model of the function field K can be projective or affine, and can also be chosen without singularities, because the singular points on a variety defined over k are a frontier F, properly contained in V and defined over k. Hence our corollaries show that as ν increases indefinitely, the number of simple points on V in k_ν also increases indefinitely.

COROLLARY 3. *Let V be an abstract variety defined over k. Let m be a given integer. There exists a zero dimensional positive cycle on V, rational over k, and of degree prime to m. Hence there exist cycles on V which are zero dimensional, rational over k, and of degree* 1.

Proof. From the above remark and the corollaries, there exist non singular points of V in the field k_ν when ν is sufficiently large and prime to m.

The preceding result generalizes the well known theorem that a function field in one variable over a finite constant field always has a divisor of degree 1. In this vein, we can state the following invariant result:

COROLLARY 4. *Let K be a function field over a finite constant field k. Let $x_1, \cdots, x_n$ be a finite set of non zero elements of K. For all ν sufficiently large, there exists a place ϕ of K which is k_ν-valued, and such that $\phi(x_i) \neq 0, \neq \infty$ for any i.*

Proof. After enlarging our set (x) by a finite number of elements if necessary, we may assume that $k(x) = K$, and that for each i, x_i^{-1} appears in the set (x). Under these assumptions it suffices to prove that there exists a place ϕ of K which is k_ν-valued and such that $\phi(x_i)$ is finite for all x_i. We view (x) as the generic point of an affine variety V defined over k. By Corollary 1, for all sufficiently large ν, there exists a point (x') of V in k_ν which is simple on V. If $\mathfrak{o}$ is the specialization ring of (x') in K and $\mathfrak{p}$ the maximal prime ideal, then $\mathfrak{o}/\mathfrak{p}$ is isomorphic to the subfield $k' = k(x')$ of k_ν. (The isomorphism is induced by the map $f(x) \to f(x')$, $f \in k[X]$). Since k is perfect, the completion of $\mathfrak{o}$ is isomorphic to a power series ring $k'\{t\} = k'\{t_1, \cdots, t_r\}$ and K can be identified with a subfield of the quotient field Ω of $k'\{t\}$. It is clear that there exists a k'-valued place ϕ of Ω mapping each t_j on 0. (For instance, view Ω as a subfield of the repeated power series field $\Gamma = k'\{t_1\}\{t_2\} \cdots \{t_r\}$. By mapping successively each t_j $(j = r, \cdots, 1)$ on 0, we get a k'-valued place ϕ of Γ.) Since the quantities x_i are in $\mathfrak{o}$ for all i, the place ϕ is finite on the x_i. The restriction of ϕ to K gives the desired place.

We now turn to the applications of Corollaries 1 and 2 to the zeta function. Let V be an abstract variety of dimension r, defined over k. We associate with V the analytic function $Z(U)$ defined by

$$d \log Z(U)/dU = \sum_{\nu=1}^{\infty} N^{(\nu)} U^{\nu-1}.$$

If V is complete and without multiple points, then $Z(U)$ is the zeta function of V [5] and Weil's conjectures state among other things that $Z(U)$ is a rational function, satisfying a functional equation of the usual type. If V has singular points, or is not complete, then the function $Z(U)$ is probably still a rational function, but it need not satisfy a functional equation.

We shall not at this point go into a detailed analysis of the relation between the functions $Z(U)$ associated with arbitrary varieties, and those associated with complete varieties without multiple points, because the results which we shall now prove as a consequence of Corollary 2 will be applicable to any variety.

COROLLARY 5. *Let V be an abstract variety of dimension r, defined over k. Then the associated function $Z(U)$ has no pole or zero in the circle*

$|U| < q^{-r}$, *and has exactly one pole of order* 1 *in the circle* $|U| < q^{-(r-\frac{1}{2})}$, *namely at* $U = q^{-r}$.

Proof. According to Corollary 2, we can write $N^{(\nu)} = q^{\nu r} + A_\nu q^{\nu(r-\frac{1}{2})}$, where A_ν is a constant, bounded in absolute value by a fixed constant A. This gives

$$d \log Z(U)/dU = q^r \sum_{\nu=0}^{\infty} (q^r U)^\nu + \sum_{\nu=1}^{\infty} A_\nu (q^{r-\frac{1}{2}} U)^\nu.$$

The first of these series is the geometric series and defines the function $1/(1-q^rU)$. Since $-q^r/(1-q^rU) = d \log (1-q^rU)/dU$ we see that $d \log [Z(U)(1-q^rU)]/dU = \sum_{\nu=1}^{\infty} A_\nu (q^{r-\frac{1}{2}}U)^\nu$. Hence $Z(U)(1-q^rU)$ has no pole or zero in the circle $|U| < q^{-(r-\frac{1}{2})}$ because the latter series converges in that circle. Furthermore q^{-r} is a pole, and is the only pole inside the circle, as was to be shown.

COROLLARY 6. *Let* V *and* V' *be two varieties which are birationally equivalent, and let* $Z(U)$ *and* $Z'(U)$ *be the associated functions. Then the function* $Z(U)/Z'(U)$ *has no zero or pole in the circle* $|U| < q^{-(r-1)}$. *Hence the zeros and poles of* $Z(U)$ *in this circle are birational invariants.*

Proof. It is clear from Corollaries 1 and 2 that the series for

$$d \log Z(U)/dU - d \log Z'(U)/dU = d \log (Z(U)/Z'(U))/dU$$

converges in the circle $|U| < q^{-(r-1)}$, and hence that $Z(U)/Z'(U)$ has no zero or pole in the circle.

Concerning the behavior of $Z(U)$ for $|U| \geqq q^{-(r-\frac{1}{2})}$ we can only make the following conjectural statements, which complement the conjectures of Weil [5].

Let P be the Picard variety of V. P is an abelian variety, whose dimension g is the irregularity of V. Let ι be the endomorphism induced on P by the automorphism $(x \to x^q)$ of the universal domain, and let

$$F(U) = \prod_{i=1}^{2g} (1 - \alpha_i U)$$

be the characteristic polynomial of ι as defined in [3], § 67. Then the series for

$$d \log Z(U)/dU + d \log (1-q^rU)/dU - d \log F(U)/dU$$

converges in the circle $|U| < q^{-(r-1)}$, and the function $Z(U)(1-q^rU)/F(U)$ has no zero and no pole inside this circle, and has at least one pole on the

circle $|U| = q^{-(r-1)}$, provided of course that V is complete and without multiple points.

Furthermore, if K/k is a function field as in Corollary 2, and if γ is the smallest constant for which Corollary 2 holds, then $\gamma = 2g$. (This gives the explicit determination of γ as a birational invariant.)

If V is a curve, then $P = J$ is the Jacobian variety, and g is the genus of the curve. For this case, all the preceding statements are well known and are contained in [2], except for the last, i. e. that $\gamma = 2g$ is actually the best possible constant for the inequality $|N^{(\nu)} - q^\nu| \leqq \gamma q^{\nu/2} + B$ of Corollary 2. As to this, it obviously suffices to prove it in the case that V is complete and without multiple points, and then it amounts to proving that $\gamma = 2g$ is the best possible constant for the inequality $|1 + q^\nu - N^{(\nu)}| \leqq \gamma q^{\nu/2}$. Actually, this last statement is implicit in [2], as the following argument shows.

We have ([2], § 22, p. 71) $|1 + q^\nu - N^{(\nu)}| = \sigma(\iota^\nu) = \sum_{i=1}^{2g} \alpha_i^\nu$ and $|\alpha_i| = q^{\frac{1}{2}}$, so that we can write $\alpha_i = q^{\frac{1}{2}} \exp(2\pi i\theta_i)$ where θ_i is a real number between 0 and 1. Hence we can write

$$\sigma(\iota^\nu) = q^{\frac{\nu}{2}} \sum_{i=1}^{2g} \exp(2\pi i\nu\theta_i).$$

For infinitely many ν, $\sum_{i=1}^{2g} \exp(2\pi i\nu\theta_i)$ will be as close to $2g$ as we please, and hence $2g$ is indeed the best possible constant.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] A. Weil, "Foundations of algebraic geometry," American Mathematical Society Colloquium Publications, Vol. XXIX, New York, 1946.

[2] ———, "Sur les courbes algébriques et les variétés qui s'en déduisent," Hermann et Cie., Paris, 1948.

[3] ———, "Variétés abéliennes et courbes algébriques," Hermann et Cie., Paris, 1948.

[4] ———, "Critères d'Equivalence," to appear in *Mathematische Annalen*.

[5] ———, "Number of solutions of equations in finite fields," *Bulletin of the American Mathematical Society*, vol. 55 (1949), no. 5, pp. 497-508.

REMARKS ON PERIODIC SEQUENCES AND THE RIEMANN ZETA-FUNCTION.*

By C. R. PUTNAM.

1. This paper is essentially a continuation of [1]. The following theorem concerning the imaginary part of the Riemann Zeta-Function $\zeta(z)$, $z = \sigma + it$, in the half-plane $\sigma > 0$, will be proved:

(*) *Let $d \neq 0$ denote a fixed real number and, for a fixed $\sigma > 0$, let $t_1, t_2, t_3, \cdots$ denote any sequence of t-values for which*

$$(1) \quad \Im(\zeta(\sigma + it_n)) - \Im(\zeta(\sigma + ind)) = o(1), \text{ as } n \to \infty.$$

Then

$$(2) \quad \limsup |\Im(\zeta(\sigma + it_n))| > 0, \text{ as } n \to \infty.$$

Furthermore, by an application of the van der Corput method for estimating certain summations, it will be shown that

(**) *If $t_n = nd + \delta_n$, then relation* (1) *(hence, by* (*)*, also* (2)*) holds, for $\frac{1}{2} \leqq \sigma < 1$, if*

$$(3) \quad \delta_n = o(n^{-1/(M-2)} \log^{-2} n), \text{ as } n \to \infty,$$

where $M = 2^m$, $\sigma = 1 - m/(M-2)$, $m \geqq 3$; and, for $\sigma = 1$, if

$$(4) \quad \delta_n = o(\log \log n / \log^2 n), \text{ as } n \to \infty.$$

The gist of (*) is that the imaginary part of $\zeta(\sigma + it)$ on any fixed line $\sigma = \text{const.} > 0$ fails to tend to the limit zero as $t = t_n \to \infty$ (or $-\infty$), whenever the sequence $t_1, t_2, t_3, \cdots$ approximates "closely enough" a periodic sequence $d, 2d, 3d, \cdots$. As a consequence of (*), one obtains the result that the sequence of the consecutive zeros of $\zeta(\frac{1}{2} + it)$ does not contain any periodic subsequence $d, 2d, 3d, \cdots$ (cf. [1]), nor even, according to (3) for $M = 8$, contain any approximately periodic sequence of the type $d + \delta_1$, $2d + \delta_2, 3d + \delta_3, \cdots$, where $\delta_n = o(n^{-1/6} \log^{-2} n)$, as $n \to \infty$.

Theorem (**) is a supplement to (*) when σ is restricted to $\frac{1}{2} \leqq \sigma \leqq 1$.

* Received December 3, 1953; revised January 28, 1954.

Moreover, estimates other than those of (3) and (4) could be obtained by other means, for instance, by the Weyl-Hardy-Littlewood method; cf. [2], pp. 84 ff.

2. *Proof of* (*). The proof of (*) will follow as a consequence of the results of [1]. Thus, in order to prove (*), it is sufficient to show that if (2) does not hold for some $d > 0$, then

$$(5) \quad \int_0^\infty g(x) \cos(dnx)\,dx = o(n^{-1}), \text{ as } n \to \infty$$

(which leads to a contradiction), where $g(x) = (e^x - n)e^{-\sigma x}$ for $\log n \leqq x < \log(n+1)$ and $n = 1, 2, \cdots$, and

$$(6) \quad \zeta(z) = z(z-1)^{-1} - z\int_0^\infty g(x)e^{-itx}dx, \quad \sigma > 0.$$

(It should be noted that there is no loss of generality in supposing $d > 0$. Moreover, although [1] dealt with the case $\sigma = \frac{1}{2}$, it is clear that any $\sigma > 0$ could be allowed.)

According to (6), for $z = \sigma + it$,

$$\int_0^\infty g(x) \cos(tx)\,dx = (\sigma-1)/((\sigma-1)^2 + t^2) - \Re(z^{-1}\zeta(z)).$$

A simple calculation shows that

$$\Re(z^{-1}\zeta(z)) = (\sigma\Re(\zeta) + t\Im(\zeta))/(\sigma^2 + t^2).$$

Since $\zeta(\sigma + it) = O(t^{\frac{1}{2}})$ as $t \to \infty$, for any fixed $\sigma > 0$ (cf., e. g., [2], p. 82), it is clear that

$$(7) \quad \int_0^\infty g(x) \cos(tx)\,dx = O(t^{-3/2}) + t^{-1}O(\Im(\zeta)).$$

Hence, if (2) fails to hold, so that

$$(8) \quad \Im(\zeta(\sigma + it_n)) = o(1), \text{ as } n \to \infty,$$

then (1) and (7) imply (5). This completes the proof of (*).

3. *Proof of* (**). It is supposed that $\frac{1}{2} \leqq \sigma < 1$, and it will be shown that (3) implies (1). Since $\zeta(\sigma + it)$ can be expressed as

$$\zeta(\sigma + it) = \sum_{n \leqq t} n^{-(\sigma+it)} + o(1), \text{ as } t \to \infty$$

(cf., e. g., [2], p. 67), an application of the mean value theorem of differ-

ential calculus shows that, for $|\delta| <$ const. and $t \to \infty$, there exist values T, lying between t and $t+\delta$, such that

$$(9) \quad \mathfrak{I}(\zeta(\sigma + i[t+\delta])) - \mathfrak{I}(\zeta(\sigma+it)) = -\delta \sum_{n \leqq t} n^{-\sigma} \log n \cos(T \log n) + o(1).$$

The summation occurring on the right side of (9) is the imaginary part of

$$(10) \quad \sum_{n \leqq t} n^{-(\sigma+iT)} \log n,$$

and so it will be sufficient to estimate the latter. Since the terms $n^{-\sigma} \log n$ constitute a monotone decreasing sequence (for n sufficiently large), the summation (10) is amenable to the van der Corput estimation procedure, just as is the summation $\sum_{n \leqq t} n^{-(\sigma+it)}$. It is easily seen that this process yields for (10), when t is large, an estimate which is $O(\log t)$ times the estimate for $\zeta(\sigma+it)$; and hence, in the notations following (3), the summation (10) is equal to $O(t^{1/(M-2)} \log^2 t)$. In this connection, see [2], pp. 94 ff. Relations (3) and (9) now imply (1).

In case $\sigma = 1$, methods similar to those described above (cf. [2], p. 97) show that (10) is equal to $O(\log^2 t / \log\log t)$ and (1) readily follows from (4) and (9). This completes the proof of (**).

PURDUE UNIVERSITY.

REFERENCES.

[1] C. R. Putnam, "On the non-periodicity of the zeros of the Riemann Zeta-Function," *American Journal of Mathematics*, vol. 76 (1954), pp. 97-99.

[2] E. C. Titchmarsh, *The Theory of the Riemann Zeta-Function*, Oxford (1951).

ON THE ESSENTIAL SPECTRA OF ORDINARY DIFFERENTIAL OPERATORS.*

By PHILIP HARTMAN.

In the differential equation

$$y'' + (\lambda + q)y = 0, \tag{1}$$

let λ be a real parameter and $q = q(t)$ a real-valued continuous function on $0 \leqq t < \infty$. When (1) is of the limit-point type (in the sense of Weyl [10]), let S' denote its essential spectrum, that is, the set of cluster points λ of the spectrum of the self-adjoint operator associated with (1) and a homogeneous boundary condition at $t = 0$. The λ-set S' is independent of the particular boundary condition determining it ([10], pp. 251-252). This note is concerned with the essential spectrum of (1) when q satisfies either the condition that

$$q(t) \text{ tends monotonously to } \infty \text{ as } t \to \infty, \tag{2}$$

or the condition that, for some constant C,

$$q(t) \leqq C \text{ for } 0 \leqq t < \infty. \tag{3}$$

1. If (2) holds, then (1) can be either of the limit-point or of the limit-circle type. Under the assumption (2), a sufficient ([6], p. 297), but not necessary ([6], pp. 306-308), condition that (1) be of the limit-point type is

$$\int^{\infty} dt/q^{\frac{1}{2}}(t) = \infty. \tag{4}$$

In [7], p. 646, there was raised the question whether or not (2) implies that

$$S' \text{ is the entire } \lambda\text{-axis} \tag{5}$$

whenever (1) is of the limit-point type. It was shown ([4], pp. 123-125) that the answer is in the negative and, in fact, that S' can be empty; but it was pointed out that there remained the question as to whether or not (2) *and* (4) imply (5), as is the case when q is of " regular growth " (for example,

* Received November 17, 1953.

when q is a logarithmico-exponential function; cf. [9], p. 106). The following theorem can be considered a partial answer to this question.

(*) *Let* $q(t)$, *where* $0 \leqq t < \infty$, *be a positive, continuous* (*non-decreasing*) *function satisfying* (2). *If, in addition,* $q(t)$ *satisfies either the condition*

$$q(t) = o(t^2) \text{ as } t \to \infty \tag{6}$$

or the condition

$$\int^{\infty} dt/q^{\frac{1}{2}\mu} = \infty \text{ for some } \mu > 1, \tag{7}$$

then (1) *is of the limit-point type and* (5) *holds.*

This theorem is known for the case when $\mu > 1$ in (7) is replaced by $\mu \geqq 2$; [6], p. 303. It will be clear from the proof of (*) that (7) can be weakened to any of the following conditions, with $\mu > 1$,

$$\text{(7 bis)} \quad \int^{\infty} dt/q^{\frac{1}{2}}(\log q)^{\mu} = \infty, \qquad \int^{\infty} dt/q^{\frac{1}{2}}(\log q)\,(\operatorname{loglog} q)^{\mu} = \infty, \cdots.$$

On the other hand, if "some $\mu > 1$" is replaced by "every $\mu < 1$" in (7) or (7 bis), then the assertion (5) need not hold. This is seen by a suitable modification of the example in [4], pp. 123-125 (thus, if $\nu_n = n^2$ is replaced by $\nu_n = n \log^2 n$ in that example, then the resulting differential equation (1) remains of the limit-point type and the corresponding S' is empty; but $q(t) = O(t^2 \log^4 t)$, so that the integral in (7) is divergent for every $\mu < 1$).

This theorem does not settle the question whether or not (2) *and* (4) imply (5). It also leaves undecided the related question whether or not (2) and the assumption that S' is not empty imply (5).

It will be clear from the proof of (*) that if the condition (6) is weakened to

$$\text{(6 bis)} \qquad q(t) \leqq Ct^2 \text{ for large } t \qquad (C = \text{Constant}),$$

then the assertion (5) of (*) is weakened to the statement that there exists a positive constant c (depending only on C and not on q) with the property that every λ-interval $[\lambda, \lambda + c]$ contains at least one point of S'.

2. In contrast to (2), the condition (3) implies that (1) is of the limit-point ([10], p. 238) and that S' is bounded from below ([10], pp. 252-253). Under the condition (3), S' can be empty (this is the case, for example, if $q(t) \to -\infty$, as $t \to \infty$; [10], p. 252). The following theorem concerning S', for the case that S' is not empty, was proved in [8]:

(**) *Let the continuous function $q = q(t)$, where $0 \leqq t < \infty$, satisfy* (3) *and let S' contain the point $\lambda = 0$. Then there exists a constant c (depending only on C and not on q) with the property that for all sufficiently large $\lambda > 0$, the interval $[\lambda, \lambda + c\lambda^{\frac{1}{3}}]$ contains at least one point of the essential spectrum S'.*

It will be shown below that this result is the "best possible" in the following sense: The length $c\lambda^{\frac{1}{3}}$ of the interval $[\lambda, \lambda + c\lambda^{\frac{1}{3}}]$ cannot be replaced by a length which (can depend on q but) is $o(\lambda^{\frac{1}{3}})$, as $\lambda \to \infty$. This will be proved by showing that

(***) *There exist continuous functions $q(t)$ satisfying* (3) *for which the essential spectrum S' of* (1) *consists of the points $\lambda = n^2$, where $n = 1, 2, \cdots$.*

In this case, there exist arbitrarily large values of λ for which the intervals $[\lambda, \lambda + 2\lambda^{\frac{1}{3}}]$ contain no points of S'.

For the sake of completeness, a simple proof of (**) will also be given below.

3. Proof of (*) when (6) is assumed. That (1) is of the limit-point type when (6), hence (4) holds, follows from [6], p. 297. Let $N(T, U; \lambda)$ denote the number of zeros on $(0 \leqq)\ T < t \leqq U$ of a (non-trivial) solution $y = y(t, \lambda)$ of (1). The assertion (5) is equivalent to the statement that

$$N(T, U; \lambda + \epsilon) - N(T, U; \lambda) \to \infty \quad \text{as} \quad U \to \infty \tag{8}$$

for every $\epsilon > 0$. For, in view of the theorem in [1], the assumption (8) is sufficient in order that the closed interval $[\lambda, \lambda + \epsilon]$, and is necessary in order that the open interval $(\lambda, \lambda + \epsilon)$, contains a point of S'.

If $\lambda + q(t) > 0$ for $t \geqq T$, then a lemma of Wintner and the author states that, for $U > T$,

$$\left| \pi N(T, U; \lambda) - \int_T^U (\lambda + q)^{\frac{1}{2}} dt \right| \leqq \pi + \tfrac{1}{4} \int_T^U | d \log(\lambda + q) | ; \tag{9}$$

[3], p. 658. Hence for fixed T, the difference in (8) exceeds

$$\int_T^U ((\lambda + \epsilon + q)^{\frac{1}{2}} - (\lambda + q)^{\frac{1}{2}}) dt - \tfrac{1}{4}\{\log(\lambda + q(U)) + \log(\lambda + \epsilon + q(U))\} + \text{const.}$$

This is minorized by $\frac{1}{2}\epsilon \int_T^U (\lambda + \epsilon + q)^{-\frac{1}{2}} dt - \frac{1}{2} \log(\lambda + \epsilon + q(U)) + \text{const.}$ The assumption (6) means that if M is an arbitrary positive constant, then

the last expression is, in turn, minorized by $M \log U - \log U + \text{Const.}$, if U is sufficiently large. This shows that (8) holds for every $\epsilon > 0$ and proves (*), when (6) is assumed.

4. Proof of (*) when (7) is assumed. Without loss of generality, it can be supposed that $q(0) = 1\ (> 0)$. It will be shown that the assumption (7) implies that if $U = U(T)$ is the largest t-value satisfying $q(t) = 2q(T)$, so that $q(U) = 2q(T)$, then

$$\limsup_{T\to\infty} \int_T^U dt/q^{\frac{1}{2}}(T) = \infty. \tag{10}$$

For if (10) does not hold, then $\int_T^U dt/q^{\frac{1}{2}}(t) \leqq \text{const.}$ Hence, for the sequence of t-values $(0 =)\ T_0 < T_1 < T_2 < \cdots$, where $T_{n+1} = U(T_n)$, so that $q(T_n) = 2^n$, one has

$$\int_0^\infty dt/q^{\frac{1}{2}\mu} < \sum_{n=0}^\infty 2^{-\frac{1}{2}(\mu-1)n} \int_u^v dt/q^{\frac{1}{2}} < \infty, \text{ where } u = T_n,\ v = T_{n+1}.$$

Since this contradicts (7), the relation (10) is proved.

If U is chosen so that $q(U) = 2q(T)$, then the monotony of q shows that (9) becomes

$$|\, \pi N(T, U; \lambda) - \int_T^U (\lambda + q)^{\frac{1}{2}} dt \,| \leqq \text{const.}$$

for large T. Hence the difference in (8) exceeds $\frac{1}{4}\epsilon \int_T^U dt/q^{\frac{1}{2}} - \text{Const.}$ for large T and $U = U(T)$. In view of (10), the limit relation (8) holds if $U = U(T)$ and T tends to ∞ through some sequence of t-values. Hence, the Sturm separation theorem shows that (8) holds (if T is fixed and U tends continuously to ∞). This proves (*) under the assumption (7).

5. Proof of (**). The assertion of (**) will be proved if it is verified that there exists a constant c (depending only on C) such that (8) holds for $\epsilon \geqq c\lambda^{\frac{1}{2}}$, whenever λ is sufficiently large.

In view of the characterization [5] of the least point of S', the assumption that $\lambda = 0$ is in S' means that (1) is oscillatory if $\lambda > 0$. Let const. > 0 and let $q + \text{const.}$ and $C + \text{const.}$ be denoted by q and C, respectively. Then (3) holds and (1) is oscillatory for $\lambda > -\text{const.}$; in particular, for $\lambda = 0$. (Incidentally, this implies $C > 0$.)

Let $y = y_0(t) \not\equiv 0$ be a solution of (1) when $\lambda = 0$ and let $t_0, t_1, \cdots$ be the sequence of successive zeros of $y_0(t)$. On an interval $t_k < t < t_{k+1}$,

the ratio $x = x(t, \lambda) = y(t, \lambda)/y_0(t)$ satisfies the differential equation $(y_0^2 x')' + \lambda y_0^2 x = 0$ and has the same zeros as $y(t, \lambda)$. If a change of independent variables $t \to s$, defined by $ds = dt/y_0^2$, is made, then the differential equation for x becomes $d^2x/ds^2 + \lambda y_0^4 x = 0$. Accordingly, the lemma of [3], mentioned in connection with (9), implies that if $\lambda > 0$ and if $u \leqq t \leqq v$ is a subinterval of $t_k < t < t_{k+1}$, then

$$| \pi N(u, v; \lambda) - \int_{s(u)}^{s(v)} \lambda^{\frac{1}{2}} y_0^2 ds | \leqq \pi + \tfrac{1}{4} \int_u^v | d \log(\lambda y_0^4) |.$$

This can be written as $| \pi N(u, v; \lambda) - (v - u)\lambda^{\frac{1}{2}} | \leqq \pi + \int_u^v | y_0'/y_0 | dt$, since $ds = dt/y_0^2$. Consequently,

$$\pi(N(u, v; \lambda + \epsilon) - N(u, v; \lambda)) \geqq (v - u)\,((\lambda + \epsilon)^{\frac{1}{2}} - \lambda^{\frac{1}{2}}) - 2\pi - 2 \int_u^v | y_0'/y_0 | dt.$$

It will be shown that the numbers $u = u_k$, $v = v_k$ can be chosen so that $u \leqq t \leqq v$ is a subinterval of $t_k < t < t_{k+1}$ and satisfies

$$v - u \geqq \tfrac{1}{2} C^{-\frac{1}{2}} \quad \text{and} \quad \int_u^v | y_0'/y_0 | dt \leqq \pi \tag{11}$$

(the point in these inequalities is that the minorant $\frac{1}{2}C^{-\frac{1}{2}}$ and the majorant π are independent of k).

Let (11) be granted for a moment. Then the last two formula lines give

$$\pi(N(u, v; \lambda + \epsilon) - N(u, v; \lambda)) \geqq \tfrac{1}{2} C^{-\frac{1}{2}}((\lambda + \epsilon)^{\frac{1}{2}} - \lambda^{\frac{1}{2}}) - 2\pi - 2\pi;$$

so that

$$N(u, v; \lambda + \epsilon) - N(u, v; \lambda) \geqq 2, \tag{12}$$

provided that $(\lambda + \epsilon)^{\frac{1}{2}} - \lambda^{\frac{1}{2}} \geqq 12\pi C^{\frac{1}{2}}$. Clearly, there exists a number $c = c(C) > 0$ such that this proviso holds if λ is sufficiently large and $\epsilon \geqq c\lambda^{\frac{1}{2}}$; in which case (12) holds. Since there is at least one zero of $y(t, \lambda + \epsilon)$ between every pair of zeros of $y(t, \lambda)$, it follows from (12) that $N(0, U; \lambda + \epsilon) - N(0, U; \lambda) \geqq k$ if $U = v_k$; and so (8) holds (with $T = 0$). Consequently (**) will be proved if the estimates (11) are established for suitably chosen u, v satisfying $t_k < u < v < t_{k+1}$.

In the proof of (11), it can be supposed that

$$y_0(t) > 0 \text{ and } \max y_0(t) = 1 \text{ for } t_k < t < t_{k+1}; \tag{13}$$

in fact, (11) and the differential equation $(y_0^2 x')' + \lambda y_0^2 x = 0$ for x show

that, on $t_k < t < t_{k+1}$, the function $y_0(t)$ can be replaced by const. $y_0(t)$, where const. can depend on k. By virtue of the lemma of [2], p. 650, it follows from (3), (13), and $y_0(t_k) = y_0(t_{k+1}) = 0$ that

$$|y_0'(t)| \leqq 2C^{\frac{1}{2}} \text{ for } t_k < t < t_{k+1}. \tag{14}$$

Choose $u = u_k$ and $v = v_k$ in such a way that

$$\tfrac{1}{2} \leqq y_0(t) \leqq 1 \text{ for } u \leqq t \leqq v \tag{15}$$

and

$$y_0(u) = y_0(v) = \tfrac{1}{2} \text{ and } y_0(w) = 1 \tag{16}$$

for some w satisfying $u < w < v$.

From (14) and (16), $\frac{1}{2} = |y_0(w) - y_0(u)| \leqq \int_u^w |y_0'(t)|\, dt$, hence $\frac{1}{2} \leqq 2C^{\frac{1}{2}}(w-u)$. Similarly, $\frac{1}{2} \leqq 2C^{\frac{1}{2}}(v-w)$, and so, by addition, $1 \leqq 2C^{\frac{1}{2}}(v-u)$. This is the first inequality in (11). Let $[u, v]$ be replaced by a subinterval (to be denoted again by $[u, v]$) such that $v - u = \frac{1}{2}C^{-\frac{1}{2}}$, so that (15) holds (but (16) need not). By (14) and (15),

$$\int_u^v |y_0'/y_0|\, dt \leqq 4C^{\frac{1}{2}}(v-u) = 2.$$

Hence, the second inequality in (11) holds. This proves (**).

6. Proof of (***). There will be constructed a step-function $q(t)$ which has discontinuities at $t = \pi, 2\pi, \cdots$, which satisfies (3) and for which the essential spectrum of (1) is the sequence of points $\lambda = 1, 2^2, 3^2, \cdots$. The passage from such a q to a continuous q with the same properties causes no difficulty; see [4], Section 11, for the treatment of a similar problem.

By a solution of (1), when q is a step-function, is meant a function $y = y(t)$ possessing, for $0 \leqq t < \infty$, a continuous first derivative and, at every continuity point of q, a continuous second derivative satisfying (1).

For $k = 0, 1, \cdots$, put

$$\begin{aligned} &q(t) = 0 \text{ for } 2k\pi \leqq t < (2k+1)\pi, \\ &q(t) = -k \text{ for } (2k+1)\pi \leqq t < (2k+2)\pi. \end{aligned} \tag{17}$$

Then

$$q(t) \leqq 0; \tag{18}$$

so that (1) is non-oscillatory for $\lambda \leqq 0$ (and, correspondingly, no negative value of λ is in S').

Let $\lambda > 0$ be fixed and $K = K(\lambda)$ so large that $K - \lambda > 0$. Then, for $t \geqq 2K\pi$, a solution $y = y(t, \lambda)$ is of the form

$$y = \alpha_k \cos \lambda^{\frac{1}{2}}(t - 2k\pi) + \beta_k \sin \lambda^{\frac{1}{2}}(t - 2k\pi),$$

$$(19)\quad y = a_k \exp\{(k-\lambda)^{\frac{1}{2}}(t - (2k+1)\pi)\} + b_k \exp\{-((k-\lambda)^{\frac{1}{2}}(t-(2k+1)\pi))\},$$

on $2k\pi \leqq t < (2k+1)\pi$, $(2k+1)\pi \leqq t < (2k+2)\pi$, respectively. The solution is determined for $t \geqq (2K+1)\pi$, for example, by assigning the constants a_K, b_K and determining α_{K+1}, β_{K+1}, a_{K+1}, b_{K+1}, $\alpha_{K+2}, \cdots$ recursively so as to saitsfy $y(t-0) = y(t+0)$ and $y'(t-0) = y'(t+0)$ at $t = (2K+1)\pi$, $(2K+2)\pi, \cdots$.

At $(2k+1)\pi$, there result the conditions

$$(20_k)\quad a_k + b_k = \alpha_k \cos + \beta_k \sin, \quad a_k - b_k = (k-\lambda)^{-\frac{1}{2}}\lambda^{\frac{1}{2}}(-\alpha_k \sin + \beta_k \cos),$$

where $\sin = \sin \lambda^{\frac{1}{2}}\pi$ and $\cos = \cos \lambda^{\frac{1}{2}}\pi$; and, at $t = (2k+2)\pi$, the conditions

$$(21_{k+1})\quad \alpha_{k+1} = a_k \exp^k + b_k \exp^{-k}, \qquad \lambda^{\frac{1}{2}}\beta_{k+1} = (k-\lambda)^{\frac{1}{2}}(a_k \exp^k - b_k \exp^{-k}),$$

where $\exp^k = \exp((k-\lambda)^{\frac{1}{2}}\pi)$ and $\exp^{-k} = 1/\exp^k$. The equations (20_k) give, as $k \to \infty$,

$$(22)\qquad a_k = \tfrac{1}{2}\alpha_k(\cos + o(1)) + \tfrac{1}{2}\beta_k(\sin + o(1)).$$

On substituting the formulae (21_k) for α_k, β_k into (22), one has

$$(23)\quad a_k = \tfrac{1}{2}a_{k-1}(1 + o(1))(\lambda^{-\frac{1}{2}}(k-1-\lambda)^{\frac{1}{2}}\sin \exp^{k-1}) + o(|\, b_{k-1}\,|),$$

if

$$(24)\qquad \sin = \sin \lambda^{\frac{1}{2}}\pi \neq 0.$$

From (21_k) and the second equation in (20_k),

$$a_k - b_k = a_{k-1}(1 + o(1))\cos \exp^{k-1} + o(|\, b_{k-1}\,|),$$

or, in view of (23),

$$(25)\qquad b_k = a_k(1 + o(1)) + o(|\, b_{k-1}\,|).$$

In (23) and (25), the $o(1)$-factors depend only on k and λ, and not on the numbers $a_K, b_K, a_{K+1}, \cdots$. Thus, if $K = K(\lambda)$ is sufficiently large and the numbers a_K and b_K, where $0 < b_K \leqq 2a_K$, determine the solution $y = y(t)$, then it follows from (23) and (25), since $(k-\lambda)^{\frac{1}{2}}\exp^k \to \infty$ as $k \to \infty$, that $|\, b_k\,| \leqq 2\,|\, a_k\,|$ and that $a_k \neq 0$ for $k = K, K+1, \cdots$. Consequently, (23) and (25) give

$$(26)\qquad a_k \sim b_k \text{ and } a_k \sim \tfrac{1}{2}a_{k-1}\lambda^{-\frac{1}{2}}(k-1-\lambda)^{\frac{1}{2}}\sin \exp^{k-1}.$$

For large k, the two numbers a_k and b_k are of the same sign; so that by the second formula in (19), $y = y(t)$ does not vanish on $(2k+1)\pi \leqq t \leqq (2k+2)\pi$. Furthermore, the first formula in (19) shows that the number of zeros on $(2k+2)\pi < t < (2k+3)\pi$ is either $[\lambda^{\frac{1}{2}}]$ or $[\lambda^{\frac{1}{2}}]+1$, where $[\lambda^{\frac{1}{2}}]$ is the greatest integer not exceeding $\lambda^{\frac{1}{2}}$. According to (26), $y((2k+1)\pi) = a_k + b_k$ and $y((2k+3)\pi) = a_{k+1} + b_{k+1}$ are of the same or of opposite sign according as $\sin \lambda^{\frac{1}{2}}\pi > 0$ or $\sin \lambda^{\frac{1}{2}}\pi < 0$. Thus the number of zeros on $(2k+1)\pi \leqq t \leqq (2k+3)\pi$, that is, on $(2k+2)\pi < t < (2k+3)\pi$, is of the same parity as $[\lambda^{\frac{1}{2}}]$ and is, therefore, equal to $[\lambda^{\frac{1}{2}}]$. Hence $N(0, U; \lambda) = U[\lambda^{\frac{1}{2}}]/2\pi + O(1)$, as $U \to \infty$.

Thus, if neither $\lambda + \epsilon$ nor λ is the square of an integer, then (8) holds (with $T = 0$) if and only if there is an integer $n > 0$ satisfying $\lambda < n^2 < \lambda + \epsilon$. In view of the characterization [1] of the points of S', the assertion (***) follows.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "A characterization of the spectra of one-dimensional wave equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 915-920.

[2] ———, "The number of L^2-solutions of $x'' + q(t)x = 0$," *ibid.*, vol. 73 (1951), pp. 635-645.

[3] ———, "On bounded Green's kernels for second order linear ordinary differential equations," *ibid.*, vol. 73 (1951), pp. 646-662.

[4] ———, "Some examples in the theory of singular boundary value problems," *ibid.*, vol. 74 (1952), pp. 107-126.

[5] ——— and C. R. Putnam, "The least cluster point of the spectra of boundary value problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[6] ——— and A. Wintner, "Criteria of non-degeneracy for the wave equation," *ibid.*, vol. 70 (1948), pp. 295-308.

[7] ——— and A. Wintner, "Oscillatory and non-oscillatory linear differential equations," *ibid.*, vol. 71 (1949), pp. 627-649.

[8] C. R. Putnam, "On the unboundedness of the essential spectrum," *ibid.*, vol. 74 (1952), pp. 578-586.

[9] E. C. Titchmarsh, *Eigenfunction Expansions*, Oxford, 1946.

[10] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

A MAXIMUM PROBLEM IN HARMONIC ANALYSIS.*

By Edwin Hewitt and Isidore Hirschman, Jr.[1]

1. Introduction. 1.1 *Mise en scène.* The classical theories of Fourier series [18] and integrals [15] have undergone a far-reaching and conceivably definitive generalization in the theory of harmonic analysis on locally compact Abelian groups. The first step in this generalization was the enunciation in 1934 of Pontryagin's duality theorem for compact Abelian groups ([10] and [17], pp. 99-109); thereafter followed rapidly the general duality theorem, Weil's generalization of the Plancherel theorem and of the Young-Hausdorff theorem, and so on ([17], pp. 111-113). A number of the more refined results of harmonic analysis on the circle and the line have nevertheless escaped attention until quite recently (see, however, [4], [5], [6], [7], and [12]).

The present paper is devoted to the extension to general locally compact Abelian groups of a classical theorem of Hardy and Littlewood dealing with Fourier series. Let p be a number such that $1 < p < 2$, and let f be a function in $\mathfrak{L}_p$ on the circle group (equivalently, let f be a measurable function on the line of period 2π such that $\int_{-\pi}^{\pi} |f(x)|^p dx < \infty$. Writing c_n for the Fourier coefficient $\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx$, $n = 0, \pm 1, \pm 2, \cdots$, one has the Hausdorff inequality

$$1.1.1 \qquad \left[\sum_{n=-\infty}^{\infty} |c_n|^{p'}\right]^{1/p'} \leq \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^p dx\right]^{1/p},$$

where $p^{-1} + p'^{-1} = 1$. (See for example [18], p. 190.) In 1926, Hardy and Littlewood [2] published a memoir in which the possibilities for equality in 1.1.1 are determined. They show (*loc. cit.* Theorem 1, p. 167) that equality occurs in 1.1.1 if and only if $f(x)$ is equal to αe^{inx} for some complex number α and some integer n. Thus, the functions in $\mathfrak{L}_p(-\pi, \pi)$ which obviously produce equality in 1.1.1 are the only functions which can do so.

The analogue of 1.1.1 for functions in $\mathfrak{L}_p(-\infty, \infty)$ is naturally a bit more subtle, since the integral

* Received February 20, 1954.

[1] This research was supported in part by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

1.1.2 $$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(x) e^{-ixy} dx$$

is undefined for functions not in $\mathfrak{L}_1(-\infty, \infty)$. The analogue of Plancherel's theorem, nevertheless, works perfectly for all $f \in \mathfrak{L}_p(-\infty, \infty)\ (1 < p < 2)$. For every such f, there exists a function f^* in $\mathfrak{L}_{p'}(-\infty, \infty)$ such that

1.1.3 $$\lim_{\substack{A \to -\infty \\ B \to \infty}} \int_{-\infty}^{\infty} \left| f^*(y) - (2\pi)^{-\frac{1}{2}} \int_A^B f(x) e^{-ixy} dx \right|^{p'} dy = 0$$

and such that

1.1.4 $$\left[(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} |f^*(y)|^{p'} dy\right]^{1/p'} \leq \left[(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} |f(x)|^p dx\right]^{1/p}.$$

This theorem is due to E. C. Titchmarsh ([14]; see also [15], pp. 96-105). A proof much simpler than those offered by Titchmarsh can be obtained from M. Riesz's convexity theorem (see [13], § 26, p. 484, or [18], p. 316, Theorem 12.41).

Diligent search and inquiry by the writers have disclosed no treatment of the case of equality in 1.1.4. It is to be conjectured, of course, that equality can occur in 1.1.4 only if $f = 0$, and we shall, in fact, show that this is so. We study here the analogue of the theorem of Hardy and Littlewood for functions which are defined on a locally compact Abelian group G and are in $\mathfrak{L}_p(G)$ for some p, $1 < p < 2$. We are able to identify all functions for which equality occurs in the analogue of 1.1.1 and 1.1.4. It is a pleasure to record our indebtedness to Professor H. S. Zuckerman for a number of valuable conversations on the subject matter of this paper; and in particular for collaboration in the construction presented in 6.5.

1.2 *Notation.* We use the following symbols throughout the present paper. R denotes the real number system, K the complex number system, T the multiplicative group of all complex numbers of absolute value 1, and N the set of all integers. For any set X and $A \subset X$, ϕ_A is defined as the function with domain X such that $\phi_A(x) = 1$ or 0 as $x \in A$ or $x \notin A$.

For a topological space X, we denote by $\mathfrak{C}_\infty(X)$ the set of all continuous complex functions f on X such that for every $\epsilon > 0$, there exists a compact subset F of X such that $|f(x)| < \epsilon$ for all $x \notin F$. The set of all bounded continuous complex functions on X will be denoted by $\mathfrak{C}(X)$.

For a measure-space X with a measure-function λ defined on some σ-algebra of subsets of X, we have the usual complex function-spaces $\mathfrak{L}_r(X)$, here considered only for $r \geq 1$. For $f \in \mathfrak{L}_r(X)$, we write, as usual,

1.2.1 $$[\int_X |f(x)|^r d\lambda(x)]^{1/r} = \|f\|_r.$$

For a topological space X and $f \in \mathfrak{C}(X)$, we write $\sup_{x \in X} |f(x)|$ as $\|f\|_\infty$.

We reserve the symbol p for numbers such that $1 < p < 2$. For every number $r \neq 1$, the number r' is defined as $r/(r-1)$.

1.3 *Hausdorff's inequality for groups.* The symbol G will be used throughout the present paper to denote an arbitrary locally compact Abelian group. The character group of G will be denoted by the symbol G^*, a character of G being taken here as a continuous homomorphism of G into or onto the group T. We write (x, y) for the value of the character $y \in G^*$ at $x \in G$, or, dually, the value of the character $x \in G$ at $y \in G^*$. The annihilator of a subset X of G is the set $N(X)$ of all $y \in G^*$ such that $(x, y) = 1$ for all $x \in X$. We select any Haar measure λ on G, and choose Haar measure μ on G^* so that equality is obtained in the Plancherel theorem. For all group-theoretic terms not explained here, the reader is referred to the monographs of A. Weil [17] or of L. H. Loomis [9]. For $f \in \mathfrak{L}_1(G)$, we write the Fourier transform

1.3.1 $$\int_G (x, y^{-1}) f(x) d\lambda(x)$$

as $f^*(y)$. For functions h in $\mathfrak{L}_1(G^*)$, the Fourier transform $h^*(x)$ of h is a function on G given by the formula

1.3.2 $$\int_{G^*} (x, y) h(y) d\mu(y).$$

Let p be a number such that $1 < p < 2$. For $f \in \mathfrak{L}_p(G)$, the Fourier transform f^* of f is defined as follows. Let $\mathfrak{S}$ be the family of simple measurable complex functions on G, that is, the measurable functions g which are in $\mathfrak{L}_1(G)$ and take on only a finite number of different values. It is clear that $\mathfrak{S} \subset \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$. For $g \in \mathfrak{S}$, we have $\|g^*\|_\infty \leq \|g\|_1$, and $\|g^*\|_2 = \|g\|_2$ by the Weil-Plancherel theorem. It is clear too that $g \in \mathfrak{L}_p(G)$. By applying M. Riesz's convexity theorem (see for example [18], p. 192 *et seq.*), one can prove without difficulty that $g^* \in \mathfrak{L}_{p'}(G^*)$ and that $\|g^*\|_{p'} \leq \|g\|_p$. Since $\mathfrak{S}$ is a dense subspace of $\mathfrak{L}_p(G)$, it is clear that the mapping $g \to g^*$ can be extended in exactly one way over $\mathfrak{L}_p(G)$ so as to be continuous and linear. Thus we can ascribe to every $f \in \mathfrak{L}_p(G)$ a Fourier transform f^* which is a function in $\mathfrak{L}_{p'}(G^*)$. It is clear that

1.3.3 $$\|f^*\|_{p'} \leq \|f\|_p$$

for all $f \in \mathfrak{L}_p(G)$, since 1. 3. 3 holds for all f in the dense subspace Σ of $\mathfrak{L}_p(G)$. We call 1. 3. 3 Hausdorff's inequality, since the proof of 1. 1. 1 for general p is due to Hausdorff. We note in passing that the image of $\mathfrak{L}_p(G)$ in $\mathfrak{L}_{p'}(G^*)$ is a proper subspace of $\mathfrak{L}_{p'}(G^*)$ for all infinite groups G (see [5]).

2. Statement of results. 2. 1 *Definition.* A function f in $\mathfrak{L}_p(G)$ for which equality holds in 1. 3. 3 is said to be a maximal function in $\mathfrak{L}_p$. A function f in $\mathfrak{L}_1(G)$ such that $\| f \|_1 = \| f^* \|_\infty$ is said to be maximal in $\mathfrak{L}_1(G)$.

2. 2 *Definition.* Let A be a compact and open subgroup of G. A function f defined on G such that

2. 2. 1 $$f(x) = \alpha(x, y)\phi_A(x),$$

where $y \in G^*$ and $\alpha \in K$, is said to be a subcharacter of the group G.

In our search for maximal functions in $\mathfrak{L}_p(G)$, we shall first show that every subcharacter and every translate of a subcharacter are maximal functions in $\mathfrak{L}_p(G)$. We make the following introductory remarks.

2. 3 THEOREM. *Let f be a function on G which is either a subcharacter or the translate of a subcharacter. Then there exist $\alpha \in K$, $x_0 \in G$, and $y_0 \in G^*$ such that*

2. 3. 1 $$f(x) = \alpha(x, y_0)\phi_{x_0 A}(x)$$

where A is a compact open subgroup of G.

This assertion is obvious.

2. 4 THEOREM. *Let f have the form* 2. 3. 1. *Then f^* has the same form on G^*, and in fact*

2. 4. 1 $$f^*(y) = \lambda(A)\alpha(x_0, y_0)(x_0^{-1}, y)\phi_{y_0 N(A)}(y),$$

$N(A)$ being a compact open subgroup of G^*.

We prove first that $N(A)$ is a compact open subgroup of G^*. It is trivial that $N(A)$ is a subgroup of G^*. Since A is compact and open, we have

2. 4. 2 $$0 < \lambda(A) < \infty.$$

The Fourier transform of ϕ_A is given by the equalities

2. 4. 3 $$\phi_A^*(y) = \int_G (x, y^{-1})\phi_A(x)\,d\lambda(x) = \int_A (x, y^{-1})\,d\lambda(x).$$

If $y \in N(A)$, it is clear that $\phi_A^*(y) = \lambda(A) > 0$. If $y \notin N(A)$, let $a \in A$ be an element for which $(a, y) \neq 1$. Then we have

2.4.4 $$\phi_A^*(y) = \int_G (ax, y^{-1})\phi_A(ax)\,d\lambda(x)$$
$$= (a, y^{-1}) \int_G (x, y^{-1})\phi_{a^{-1}A}(x)\,d\lambda(x) = (a, y^{-1})\phi_A^*(y).$$

The first and last terms of 2.4.4 can be equal only if $\phi_A^*(y) = 0$, and therefore we find

2.4.5 $$\phi_A^* = \lambda(A)\phi_{N(A)}.$$

Now ϕ_A^* is continuous, and hence $N(A)$ is open and closed. Furthermore, ϕ_A^* is in $\mathfrak{C}_\infty(G^*)$, and this implies that $N(A)$ is compact. Returning to our original f, we can now easily compute f^*. In fact:

2.4.6 $$f^*(y) = \alpha \int_G (x, y^{-1})(x, y_0)\phi_{x_0A}(x)\,d\lambda(x)$$
$$= \alpha \int_G (x, y^{-1})(x, y_0)\phi_A(x_0^{-1}x)\,d\lambda(x)$$
$$= \alpha \int_G (x_0x, y^{-1})(x_0x, y_0)\phi_A(x)\,d\lambda(x)$$
$$= \alpha(x_0, y_0)(x_0, y^{-1}) \int_G (x, y_0y^{-1})\phi_A(x)\,d\lambda(x).$$

By 2.4.5, the integral appearing in the preceding line is equal to $\lambda(A)\phi_{N(A)}(y_0^{-1}y) = \lambda(A)\phi_{y_0N(A)}(y)$, and this completes the proof.

2.5 THEOREM. *Every subcharacter and every translate of a subcharacter of G are elements of $\mathfrak{L}_p(G)$ for all p, $1 < p < 2$, and are maximal functions in every $\mathfrak{L}_p(G)$.*

Let f be a function of the form 2.3.1. We lose no generality by taking $\alpha = 1$ in 2.3.1. Then we have

2.5.1 $$\| f \|_p = [\int_{x_0A} 1\,d\lambda(x)]^{1/p} = \lambda(A)^{1/p},$$

and

2.5.2 $$\| f \|_2 = [\lambda(A)]^{\frac{1}{2}}.$$

For every $r \geq 1$, we have

2.5.3 $$\| f^* \|_r = \lambda(A)[\int_{y_0N(A)} 1\,d\mu(y)]^{1/r} = \lambda(A)[\mu(N(A))]^{1/r}.$$

By the Weil-Plancherel theorem and our choice of the Haar measure μ on G^*, we have

2.5.4 $$[\lambda(A)]^{\frac{1}{2}} = \| f \|_2 = \| f^* \|_2 = \lambda(A)[\mu(N(A))]^{\frac{1}{2}},$$

upon equating 2.5.2 and 2.5.3 with $r = 2$. 2.5.4 implies that

2.5.5 $$\mu(N(A)) = [\lambda(A)]^{-1}$$

Hence $\| f^* \|_{p'} = \lambda(A)\,[\lambda(A)]^{-1/p'} = [\lambda(A)]^{1/p} = \| f \|_p$.

Theorem 2.5 indicates that subcharacters of G and their translates play the same role for general locally compact Abelian groups that characters of T do for the theorem of Hardy and Littlewood. That is, they are functions in $\mathfrak{L}_p(G)$ which are obviously maximal. As in the case of the theorem of Hardy and Littlewood, we now prove that subcharacters and their translates are the *only* maximal functions. Our results are comprised in the following theorem.

2.6 THEOREM. *Let G be any locally compact Abelian group and let p be a number such that $1 < p < 2$. Then if f is a maximal function in $\mathfrak{L}_p(G)$, f is either a subcharacter or a translate of a subcharacter.*

2.7 *Remark.* Theorem 2.6 may be regarded as a direct generalization of the theorem of Hardy and Littlewood, since the group T obviously has no compact open subgroups other than itself. Since R has no compact open subgroups at all, Theorem 2.6 implies that no function $\neq 0$ in $\mathfrak{L}_p(R)$ can be maximal.

2.8 *Method of proof.* The fundamental inequality 1.3.3 is proved by using the Riesz convexity theorem. Our analysis of the case of equality, which occupies us here, depends upon obtaining conditions for equality in Riesz's theorem, and exploiting the consequences of these conditions. In this process, the conditions for equality in Hölder's inequality play an important role.

3. Equality in Hölder's inequality. Throughout this Section, G is an arbitrary locally compact Abelian group. The following theorem, while a consequence of well-known facts, does not seem to appear in the literature. We take $1' = \infty$ and $\mathfrak{L}_\infty(G)$ as usual the space of essentially bounded measurable functions on G with $\| g \|_\infty = \operatorname{ess\,sup} | g |$.

3.1 THEOREM. *Let $1 \leq r < \infty$. Let $f \in \mathfrak{L}_r(G)$ and $g \in \mathfrak{L}_{r'}(G)$ and suppose that f is different from 0 on a set of positive measure. The relation*

3.1.1 $$\int_G f(x)g(x)\,d\lambda(x) = \| f \|_r \cdot \| g \|_{r'} \neq 0$$

holds if and only if

3.1.2 $g(x) = \| g \|_{r'} \| f \|_r^{1-r} |f(x)|^{r-1} \operatorname{sgn} \tilde{f}(x)$ a. e. on $G (1 < r < \infty)$;

3.1.3 $g(x) = \| g \|_\infty \operatorname{sgn} \tilde{f}(x)$ a. e. on the set where $f \neq 0$ $(r = 1)$.

(For $z \in K$, $\operatorname{sgn} z$ is defined as $z/|z|$ if $z \neq 0$, and as 0 if $z = 0$.) For $f(x)$ and $g(x)$ real and for $1 < r < \infty$, this result is demonstrated in [3, p. 140]. Only very minor changes are needed to extend the proof given there so that it applies to complex functions and to the case $r = 1$.

4. Some preliminary results. Let G be, as usual, a locally compact Abelian group. It follows immediately from Theorem 3.1 that a function $f \in \mathfrak{L}_1(G)$ is maximal if and only if $f(x) = c(y, x)p(x)$ a. e. in G, where c is a complex constant, y is an element of G^*, and p is a non-negative function in $\mathfrak{L}_1(G)$. Further information concerning this case is contained in the following assertion.

4.1 THEOREM. *Let f be a maximal function in $\mathfrak{L}_1(G)$. Let E^* be the set of points y for which $|f^*(y)| = \| f \|_1$. Then $E^* = y_0 E_0^*$, where E_0^* is a compact subgroup of G^* and $y_0 \in G^*$.*

Let A be the set of points x for which $f(x) \neq 0$. Then $y \in E^*$ if and only if $\operatorname{sgn} [(y, x) f(x)]$ is constant almost everywhere in A, as Theorem 3.1 shows. Let $y_0 \in E^*$, so that $\operatorname{sgn} [(y_0, x) f(x)] = c_0$, $|c_0| = 1$ a. e. in A. We set $E_0^* = y_0^{-1} E^*$. Let $y_1, y_2 \in E_0^*$; then $\operatorname{sgn} [(x, y_0)(x, y_1) f(x)] = c_1$, $|c_1| = 1$ a. e. in A, and $\operatorname{sgn} [(x, y_0)(x, y_2) f(x)] = c_2$, $|c_2| = 1$ a. e. in A. A simple computation now shows that $\operatorname{sgn} [(x, y_0)(x, y_1 y_2) f(x)] = c_0^{-1} c_1 c_2$ a. e. in A, *i. e.*, $y_1 y_2 \in E_0^*$. Again, if $y_1 \in E_0^*$, then $\operatorname{sgn} [(x, y_0)(x, y_1^{-1}) f(x)] = c_0^2 c_1^{-1}$ a. e. in A, *i. e.*, $y_1^{-1} \in E_0^*$. Thus E_0^* is a subgroup of G^*. If k is any function in $\mathfrak{L}_1(G)$, if $a > 0$, and if K^* is the set of points y in G^* for which $|[k^*(y)]| \geqq a$, then K^* is compact, since $k^* \in \mathfrak{C}_\infty(G^*)$. In particular E^*, and therefore E_0^*, are compact.

We shall also require the following result.

4.2 THEOREM. *Let f be a maximal function in $\mathfrak{L}_p(G)$. Then the function $h = |f^*|^{p'-2} \overline{f^*}$ is maximal in $\mathfrak{L}_p(G^*)$, and its Fourier transform (1.3.2) in $\mathfrak{L}_{p'}(G)$ is equal to*

4.2.1 $$\| f \|_p^{p'-p} |f|^{p-2} \tilde{f},$$

this function being taken as 0 wherever $f = 0$.

As f is maximal, we have by definition $\| f \|_p = \| f^* \|_{p'}$. The function $h = |f^*|^{p'-2} \tilde{f}^*$ is in $\mathfrak{L}_p(G^*)$, since $|h|^p = |f^*|^{(p'-1)p} = |f^*|^{p'}$. We have

furthermore $\| h \|_p = \| f^* \|_{p'}^{1/(p-1)} = \| f \|_p^{1/(p-1)} = \| f \|_p^{p'/p}$. For each $g \in \mathfrak{L}_p(G)$, consider the number

$$H(g) = \int_{G^*} g^*(y) h(y) \, d\mu(y).$$

It is clear that H is a bounded linear functional on $\mathfrak{L}_p(G)$ and that

4.2.2 $$\| H \| = \sup_{\|g\|_p \leq 1} |H(g)| \leq \{ \sup_{\|g\|_p \leq 1} \| g^* \|_{p'} \} \| h \|_p = \| h \|_p = \| f \|_p^{p'/p}.$$

Taking $g = \| f \|_p^{-1} f$, we find

4.2.3 $$\| H \| \geq \int_{G^*} \| f \|_p^{-1} f^*(y) h(y) \, d\mu(y) = \| f \|_p^{-1} \| f^* \|_{p'}^{p'}$$
$$= \| f \|_p^{p'-1} = \| f \|_p^{1/(p-1)}.$$

Combining 4.2.2 and 4.2.3, we find that $\| H \| = \| f \|_p^{1/(p-1)}$. It follows from the representation theorem for linear functionals on $\mathfrak{L}_p$ (see for example [9], p. 41) that there is a function η in $\mathfrak{L}_{p'}(G)$ such that $H(g) = \int_G g(x) \eta(x) \, d\lambda(x)$ for all $g \in \mathfrak{L}_p(G)$; *i. e.*,

4.2.4 $$\int_{G^*} g^*(y) h(y) \, d\mu(y) = \int_G g(x) \eta(x) \, d\lambda(x).$$

We know too that $\| \eta \|_{p'} = \| H \| = \| f \|_p^{1/(p-1)}$.

To identify the function η, put $g = f$ in 4.2.4; this produces the inequalities

4.2.5 $$\int_G f(x) \eta(x) \, d\lambda(x) = \int_{G^*} f^*(y) h(y) \, d\mu(y) = \| f^* \|_{p'}^{p'} = \| f \|_p^{1+p'-1}$$
$$= \| f \|_p \| f \|_p^{1/(p-1)} = \| f \|_p \| \eta \|_{p'}.$$

Comparing the initial and final terms in 4.2.5 and applying Theorem 3.1, we infer that

4.2.6 $$\eta(x) = \| \eta \|_{p'} \| f \|_p^{1-p} | f(x) |^{p-2} \bar{f}(x) = \| f \|_p^{p'-p} | f(x) |^{p-2} \bar{f}(x)$$

for almost all $x \in G$ (the last function being taken as 0 wherever $f(x) = 0$). Finally, the relations 4.2.4 suffice to show that η is the Fourier transform of h (see [17], p. 118). This establishes the present theorem, as $\| \eta \|_{p'}$ and $\| h \|_p$ are evidently equal.

5. Proof of Theorem 2.6. 5.1 If $f \in \mathfrak{L}_p(G)$ or $g \in \mathfrak{L}_p(G^*)$, we set $T(w) f = | f |^{p(w+1)/2} \operatorname{sgn} f$ and $T(w) g = | g |^{p(w+1)/2} \operatorname{sgn} g$. Note that if $q(w) = 2/(u+1)$, where $w = u + iv$, then $\| T(w) f \|_{q(w)}^{q(w)} = \| f \|_p^p$,

$\| T(w)g \|_{q(w)}{}^{q(w)} = \| g \|_{p}{}^{p}$. Furthermore, $q(0+iv) = 2$ and $q(1+iv) = 1$, $-\infty < v < \infty$, and $T(-1+2/p)f = f$, $T(-1+2/p)g = g$.

Let $I_1, \cdots, I_N$ be disjoint measurable sets in G, each of finite measure. Let

$$F(x) = \begin{cases} \exp(A_n + iB_n) & x \in I_n;\ A_n,\ B_n \in R; \qquad (n = 1, \cdots, N) \\ 0 & x \notin \bigcup_{n=1}^{N} I_n. \end{cases}$$

Similarly, let $J^*_1, \cdots, J^*_M$ be disjoint measurable sets in G^*, each of finite measure, and let

$$H(y) = \begin{cases} \exp(C_m + iD_m) & y \in J^*_m;\ C_m,\ D_m \in R; \quad (m = 1, \cdots, M). \\ 0 & y \notin \bigcup_{m=1}^{M} J^*_m. \end{cases}$$

Function of the type of F and H are called simple functions. Consider

$$\Psi(w) = \int_{G^*} [T(w)F]^*(y)\,[T(w)\bar{H}(y)]\,d\mu(y).$$

By direct computation, we have

$$\Psi(w) = \sum_{n=1}^{N} \sum_{m=1}^{M} \exp[A_n p(w+1)\tfrac{1}{2} + iB_n + C_m p(w+1)\tfrac{1}{2} - iD_m] \cdot \int_{I_n} \int_{J^*_m} (x, y)\,d\mu(y)\,d\lambda(x).$$

It follows that $\Psi(w)$ is an entire function.

Let f be a maximal function in $\mathfrak{L}_p(G)$. Let $f^{\#} = |f^*|^{p'-1} \operatorname{sgn} f^*$. Choose a sequence of simple functions F_α $(\alpha = 1, 2, \cdots)$ on G such that

5.1.1 $$\| T(w)F_\alpha \|_{q(w)} \leq \| T(w)f \|_{q(w)} \qquad (0 \leq \Re w \leq 1),$$

5.1.2 $$\lim_{\alpha \to \infty} \| T(w)F_\alpha - T(w)f \|_{q(w)} = 0,$$

uniformly for w in any compact subset of $0 \leq \Re w \leq 1$. Similarly choose a sequence of simple functions H_α on G^* $(\alpha = 1, 2, \cdots)$ such that

5.1.3 $$\| T(w)H_\alpha \|_{q(w)} \leq \| T(w)f^{\#} \|_{q(w)} \qquad (0 \leq \Re w \leq 1),$$

5.1.4 $$\lim_{\alpha \to \infty} \| T(w)H_\alpha - T(w)f^{\#} \|_{q(w)} = 0$$

uniformly for w in any compact subset of $0 \leq \Re w \leq 1$. It is evident that this can be done. If

$$\psi_\alpha(w) = \int_{G^*} [T(w)F_\alpha]^*(y)\,[T(w)\bar{H}_\alpha(y)]\,d\mu(y),$$

then for each α, $\psi_\alpha(w)$ is an entire function of w. Moreover, it follows from

5. 1. 2 and 5. 1. 4 that $\lim \psi_\alpha(w) = \psi(w)$, as $\alpha \to \infty$, uniformly for w in any compact subset of $0 \leq \Re w \leq 1$, where

5. 1. 5 $$\psi(w) = \int_{G^*} [T(w)f]^*(y)\,[T(w)\overline{f^\#}(y)]\,d\mu(y).$$

Thus we see that

a. $\psi(w)$ is analytic for $0 < \Re w < 1$ and is continuous for $0 \leq \Re w \leq 1$.

It follows from Hölder's inequality that

$$|\psi_\alpha(w)| \leq \|[T(w)F_\alpha]^*\|_{q(w)'} \|T(\bar{w})H_\alpha\|_{q(w)}.$$

From 5. 1. 4 and Hausdorff's inequality 1. 3. 3, we have

$$\|T(\bar{w})H_\alpha\|_{q(w)} \leq \|T(\bar{w})f^\#\|_{q(w)} = \|f^\#\|_p^{p/q(w)} = \|f^*\|_{p'}^{p'/q(w)} = \|f\|_p^{p'/q(w)}.$$

From 5. 1. 2 and 1. 3. 3, we obtain

$$\|[T(w)F_\alpha]^*\|_{q(w)'} \leq \|T(w)F_\alpha\|_{q(w)} \leq \|T(w)f\|_{q(w)} = \|f\|_p^{p/q(w)}.$$

Thus $|\psi_\alpha(w)| \leq \|f(x)\|_p^{(p+p')/q(w)}$ $(\alpha = 1, 2, \cdots; 0 \leq \Re w \leq 1)$. Let us suppose, for the sake of simplicity, that $\|f\|_p = 1$. Then we have $|\psi_\alpha(w)| \leq 1$ $(\alpha = 1, 2, \cdots; 0 \leq \Re w \leq 1)$, and thus

b. $|\psi(w)| \leq 1$ $\qquad 0 \leq \Re w \leq 1.$

Finally, since f is maximal in $\mathfrak{L}_p(G)$,

c. $\psi(-1 + 2/p) = 1.$

The statements a., b. and c. together with the maximum modulus principle imply that

5. 1. 6 $$\psi(w) \equiv 1 \qquad 0 \leq \Re w \leq 1.$$

If we estimate $|\psi(w)|$ using Hölder's inequality just as we previously estimated $|\psi_\alpha(w)|$, we find that this estimate is 1. Thus we may apply Theorem 3. 1 to conclude that

5. 1. 7 $$[T(w)f]^*(y) = |T(\bar{w})f^\#(y)|^{q(w)-1} \operatorname{sgn} [T(\bar{w})f^\#(y)] \quad (0 \leq \Re w < 1)$$

for a. a. $y \in G^*$. (The exceptional set can of course depend on w). For $\Re w = 1$, we find that if E_w^* is defined as the set of $y \in G^*$ for which $|[T(w)f]^*(y)| = 1$, then

5. 1. 8 $$T(\bar{w})f^*(y) = 0 \qquad (\Re w = 1)$$

for almost all $y \notin E_w^*$ (again the exceptional set may depend on w).

In Theorem 4. 2, it was shown that $f^\#$ is maximal in $\mathfrak{L}_p(G^*)$ and that

$$\int_{G^*} f^\#(y)\,(y, x)\,d\mu(y) = |f(x)|^{p-1} \operatorname{sgn} f(x).$$

From this it follows that f plays the same rôle with respect to $f^\#$ that $f^\#$ does with respect to f. Consequently, we may conclude that the Fourier transform on G of $T(w)f^\#$ is given by the expression

5. 1. 7′ $$[T(w)f(x)]^{q(w)-1} \operatorname{sgn}[T(w)f(x)] \qquad (0 \leq \Re e\, w < 1)$$

almost everywhere. For $\Re e\, w = 1$, we find that if $D_w \subset G$ is defined as the set of points where the Fourier transform on G of $T(w)f^\#$ has absolute value 1, then

5. 1. 8′ $$|T(w)f(x)| = 0 \qquad (\Re e\, w = 1)$$

for almost all $x \notin D_w$.

5. 2 Taking $w = 1$ in 5. 1. 8 and 5. 1. 8′ and using Theorem 4. 1, we see that f vanishes almost everywhere except on the compact set D_1 and $f^\#(x)$ vanishes almost everywhere except on the compact set $E_1{}^*$. We immediately infer from this that f and $f^\#$ may be taken as continuous and that 5. 1. 7 and 5. 1. 7′ then obtain everywhere. Consequently, we may now permit w to approach $1 + iv$ in 5. 1. 7 and 5. 1. 7′ to obtain the following equalities:

5. 2. 1 $$[T(1+iv)f]^*(y) = \operatorname{sgn}[T(1-iv)f^\#(y)], \quad y \in G^*,$$

5. 2. 1′ $$[T(1+iv)f^\#]^*(x) = \operatorname{sgn}[T(1-iv)f(x)], \quad x \in G.$$

Taking $v = 0$, we see that

5. 2. 2 $$f^\#(y) = 0 \quad \text{for} \quad y \notin E_1{}^*;\ f^\#(y) \neq 0 \quad \text{for} \quad y \in E_1{}^*,$$

5. 2. 2′ $$f(x) = 0 \quad \text{for} \quad x \notin D_1;\ f(x) \neq 0 \quad \text{for} \quad x \in D_1.$$

Since the complement of $E_1{}^*$ is exactly the set where $f^\#$ vanishes, it is closed, and $E_1{}^*$ is open. Thus $E_1{}^* = y_0 E_0{}^*$, where $E_0{}^*$ is an open compact subgroup of G. Similarly, $D_1 = x_0 D_0$, where D_0 is an open compact subgroup of G.

It follows from 5. 2. 1 that for all $v\,(-\infty < v < \infty)$, $|[T(1+iv)f]^*(y)| = 1$, $y \in E_1{}^*$. Taking an arbitrary $y_1 \in E_1{}^*$, we must have

5. 2. 3 $$\operatorname{sgn}[T(1+iv)f(x)\,(x, y_1{}^{-1})] = d_v, \quad x \in D_1,$$

where d_v is a constant of absolute value 1, independent of x. Rewriting 5. 2. 3, we have $|f(x)|^{ipv/2} \operatorname{sgn}[f(x)\,(x, y_1{}^{-1})] = d_v$, $x \in D_1$. Let $v = 0$; we find that $\operatorname{sgn} f(x) = (x, y_1)\,d_0$, $x \in D_1$. Consequently, if $-\infty < v < \infty$, then $|f(x)|^{ipv/2} = d_v/d_0$, $x \in D_1$. This is possible only if $|f(x)|$ is a constant, say

k, throughout D_1. Thus we have proved that f has the form $f(x) = kd_0(x, y_1)$ or $f(x) = 0$ according as $x \in x_0 D_0$ or $x \notin x_0 D_0$, *i. e.*, f is a subcharacter or a translate of a subcharacter.

6. Examples. 6. 1 If G is any compact connected Abelian group, then the only maximal functions in $\mathfrak{L}_p(G)$, for any p such thot $1 < p < 2$, are multiples of characters. This proves the original theorem of Hardy and Littlewood ($G = T$). Theorem 2. 6 shows this to be the case, since G can have no proper open and closed subgroups.

6. 2 If G is any connected, locally compact, and non-compact Abelian group (for example, R or R^n $(n = 2, 3, \cdots)$), then $\mathfrak{L}_p(G)$ contains no maximal functions at all save 0, since G has no compact open subgroups.

6. 3 A discrete Abelian group G always admits maximal functions: any function vanishing except at a single point in G is maximal in every $\mathfrak{L}_p(G)$ $(1 < p < 2)$. Non-trivial maximal functions exist if and only if G has finite proper subgroups, and then they are completely identified by Theorem 2. 6.

6. 4 There exist, of course, non-trivial examples of locally compact Abelian groups having compact open subgroups. As Pontryagin has proved ([11], p. 77, Theorem 17), if G is locally compact, Abelian, 0-dimensional, and has a countable open basis, then it has arbitrarily small (hence compact) open-and-closed subgroups. All such groups admit non-trivial maximal functions in every $\mathfrak{L}_p(G)$, as Theorem 2. 5 shows.

6. 5 An intriguing example of the situation described in 6. 4 is presented by the p-adic numbers Ω_p under addition. We represent Ω_p as the group of all formal power series

$$\boldsymbol{a} = \sum_{k=n}^{\infty} a_k p^k,$$

where n is an arbitrary integer and a_k can be any of the numbers $0, 1, 2, \cdots, p - 1$. Addition of two p-adic numbers $\boldsymbol{a}$ and $\boldsymbol{b}$ is carried out by adding the terms a_k and b_k and carrying remainders modulo p in the usual way. A generic neighborhood Γ_m of zero consists of all $\boldsymbol{a}$ such that $a_k = 0$ for $k < m$; here m is an arbitrary integer. Every Γ_m is clearly a compact open subgroup of Ω_p. Haar measure on Ω_p is completely determined by specifying the measure λ for all of the sets Γ_m. (It is easy to see that this fixes λ for all compact subsets of Ω_p and thence, *via* the usual construction, for all Borel sets in Ω_p.) As Γ_0 is the compact group of p-adic integers, we deem it appropriate to take $\lambda(\Gamma_0) = 1$. Then a simple calculation shows that $\lambda(\Gamma_m) = p^{-m}$

for all integers m. The character group Ω_p^* can be constructed as follows (see [8], pp. 564-565). A generic character of Ω_p is a function γ_{nh} such that

$$6.5.1 \qquad \gamma_{nh}(\boldsymbol{a}) = \gamma_{nh}\Big(\sum_{k=m}^{\infty} a_k p^k\Big) = \exp\Big[2\pi i p^{-n} h\Big(\sum_{k=-\infty}^{\infty} a_k p^k\Big)\Big]$$

$$= \exp\Big[2\pi i p^{-n} h\Big(\sum_{k=m}^{n-1} a_k p^k\Big)\Big].$$

Here n and h are integers such that $p \nmid h$. It is not hard to see that a complete family of neighborhoods of the identity γ_{00} in Ω_p^* are the subgroups $N(\Gamma_m)$ $(m = 0, \pm 1, \pm 2, \cdots)$; note that $N(\Gamma_m)$ is the set of all γ_{nh} such that $n \leq m$, and that as m decreases, Γ_m increases and $N(\Gamma_m)$ decreases. As shown in 2.5.5, we have $\mu(N(\Gamma_m)) = p^m$ for all integers m.

6.6 Finally, we consider the group D which is the direct product of a countably infinite number of groups each consisting of exactly two elements. Each factor group may be taken as the numbers -1, 1 under multiplication, and thus D is representable as the group of all sequences $\boldsymbol{x} = \{x_n\}_{n=1}^{\infty}$ in which each x_n is either -1 or 1 and the group operation is co-ordinate-wise multiplication. The neighborhood $U_m(e)$ $(m = 1, 2, 3, \cdots)$ consists of all $\boldsymbol{x}$ such that $x_1 = x_2 = x_3 = \cdots = x_m = 1$, and these neighborhoods are a complete family of neighborhoods of the identity. Each U_m is an open-and-closed subgroup of D. There are of course other such subgroups: they are all identified as the sets of $\boldsymbol{x} \in D$ for which a certain specified finite set of co-ordinates are all equal to one. It is furthermore easy to see that a translate of an open-and-closed subgroup of D consists of all $\boldsymbol{x} \in D$ for which x_i has prescribed values at a fixed, finite, number of co-ordinates. Finally, one verifies without difficulty that a maximal function in $\mathfrak{L}_p(D)$ is any linear combination of characters of D which has constant absolute value on one of these translates and is zero elsewhere. (Characters of D are just the functions $\boldsymbol{x} \to x_{i_1} x_{i_2} \cdots x_{i_m}$, where $\{i_1, i_2, \cdots, i_m\}$ is any finite set of positive integers.) It was a close study of certain maximal functions in $\mathfrak{L}_p(D)$ that led to the formulation of Theorem 2.6.

THE UNIVERSITY OF WASHINGTON
AND
WASHINGTON UNIVERSITY.

BIBLIOGRAPHY.

[1] Cartan, Henri, and Godement, Roger, "Théorie de la dualité et analyse harmonique dans les groupes abéliens localement compacts," *Annales de l'École Normale Superieur*, Ser. 3, vol. 64 (1947), pp. 79-99.

[2] Hardy, G. H., and Littlewood, J. E., "Some new properties of Fourier constants," *Mathematische Annalen*, vol. 97 (1926), pp. 158-209.

[3] ———, ———, and Polyá, G., *Inequalities*, 2nd ed., Cambridge University Press, 1952.

[4] Helson, Henry, "Spectral synthesis of bounded functions," *Arkiv för Mat.*, vol. 1 (1951), pp. 497-502.

[5] Hewitt, Edwin, "Fourier transforms of the class $\mathfrak{L}_p$," ibid., vol. 2 (1954), pp. 571-574.

[6] ———, "Linear functionals on almost periodic functions," *Transactions of the American Mathematical Society*, vol. 74 (1953), pp. 303-322.

[7] ———, "Remarks on the inversion of Fourier-Stieltjes transforms," *Annals of Mathematics*, vol. 57, ser. 2 (1953), pp. 458-474.

[8] ———, and Zuckerman, H. S., "A group-theoretic method in approximation theory," *ibid.*, vol. 52, ser. 2 (1950), pp. 557-567.

[9] Loomis, Lynn H., *An introduction to abstract harmonic analysis*, D. van Nostrand Company, Inc., New York, 1953.

[10] Pontryagin, L. S., "The theory of topological commutative groups," *Annals of Mathematics*, vol. 35, ser. 2 (1934), pp. 361-388.

[11] ———, *Topological groups*, Princeton University Press, 1939.

[12] Reiter, Hans J., "Investigations in harmonic analysis," *Transactions of the American Mathematical Society*, vol. 73 (1952), pp. 401-427.

[13] Reisz, M., "Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires," *Acta Mathematica*, vol. 49 (1926), pp. 465-497.

[14] Titchmarsh, E. C., "A contribution to the theory of Fourier transforms," *Proceedings of the London Mathematical Society*, vol. 23, ser. 2 (1923), pp. 279-289.

[15] ———, *Introduction to the theory of Fourier integrals*, Oxford University Press, 1937.

[16] Verblunsky, S., "Fourier constants and Lebesgue classes," *Proceedings of the London Mathematical Society*, vol. 39, ser. 2 (1935), pp. 1-31.

[17] Weil, André, *L'intégration dans les groupes topologiques et ses applications*, Hermann et Cie., Paris, 1940.

[18] Zygmund, Antoni, *Trigonometrical series*, Monografje Matematyczne. Warszawa-Lwow, 1935.

ALGEBRAS WITH TWO GENERATORS.*

By JOHN WERMER.

Introduction. Let $\mathbf{C}$ denote the algebra of all continuous complex-valued functions defined on the unit circle $|\lambda| = 1$. With the norm $\|f\| = \sup_{|\lambda|=1} |f(\lambda)|$, $\boldsymbol{C}$ is a Banach algebra.

For any ϕ, ψ in $\mathbf{C}$, let $[\phi, \psi]$ be the closed subalgebra of $\mathbf{C}$ generated by ϕ, ψ and the constant 1. $[\phi, \psi]$ then consists of those functions in $\mathbf{C}$ which can be uniformly approximated by polynomials in ϕ and ψ.

In [1] the author determined the algebras $[\phi, \psi]$ with ϕ one-one on $|\lambda| = 1$. In this paper we shall give explicitly all algebras $[\phi, \psi]$ where $\phi(\lambda) = \lambda^2$ and ϕ and ψ together separate points on $|\lambda| = 1$. In § 2 we discuss the general problem: when is $[\phi, \psi]$ equal to $\mathbf{C}$?

1. We assume λ^2 and ψ separate points on $|\lambda| = 1$. We shall prove:

THEOREM 1. *If $[\lambda^2, \psi] \neq \mathbf{C}$, then there exist n distinct points $\lambda_1^2, \lambda_2^2, \cdots, \lambda_n^2$ in $|\lambda| < 1$, where n is odd, and functions E_1, E_2 analytic in $|\lambda| < 1$ and continuous in $|\lambda| \leqq 1$, such that $\psi(\lambda) = E_1(\lambda^2) + \{\prod_{i=1}^{n} (\lambda^2 - \lambda_i^2)\}^{\frac{1}{2}} E_2(\lambda^2)$. Conversely, if ψ has this form, then $[\lambda^2, \psi] \neq \mathbf{C}$.*

Definition. Let $\mathbf{R}$ be any Banach algebra, $\mathbf{M}$ a closed subalgebra. We say $\mathbf{M}$ is a *maximal* subalgebra of $\mathbf{R}$ if for any closed subalgebra $\mathbf{M}'$ of $\mathbf{R}$ with $\mathbf{M} \subseteq \mathbf{M}'$ we either have $\mathbf{M}' = \mathbf{M}$ or $\mathbf{M}' = \mathbf{R}$.

THEOREM 2. *Every subalgebra $[\lambda^2, p]$, where $p(\lambda) = \{\prod_{i=1}^{n} (\lambda^2 - \lambda_i^2)\}^{\frac{1}{2}}$, the λ_i are distinct points in $|\lambda| < 1$ and n is odd, is a maximal subalgebra of C.*

We shall use the following notations: If $h(\lambda)$ is any function defined on $|\lambda| = 1$, then $h_0(\lambda) = \frac{1}{2}(h(\lambda) - h(-\lambda))$, $h_e(\lambda) = \frac{1}{2}(h(\lambda) + h(-\lambda))$. Clearly $h = h_e + h_0$.

If S is any set on $|\lambda| = 1$, $-S = \{\lambda \mid -\lambda \in S\}$.

If μ is any complex-valued measure on $|\lambda| = 1$, then

* Received March 1, 1954.

$$\mu_0(S) = \tfrac{1}{2}(\mu(S) - \mu(-S)), \qquad \mu_e(S) = \tfrac{1}{2}(\mu(S) + \mu(-S)).$$

Clearly μ_0 and μ_e again are measures on $|\lambda| = 1$, and $\mu = \mu_e + \mu_0$.

$\mathfrak{A}$ denotes the class of functions $f(\lambda)$ in $\mathbf{C}$ with $f(\lambda)$ analytic in $|\lambda| < 1$, continuous in $|\lambda| \leqq 1$.

$\mathfrak{H}'$ denotes the class of function $h(\lambda)$ on $|\lambda| = 1$ such that there exists $H(\lambda)$ analytic in $|\lambda| < 1$ with $\sup_{r<1} \int_0^{2\pi} |H(re^{i\theta})|\, d\theta < \infty$ and $h(e^{i\theta}) = \lim_{r\to 1} H(re^{i\theta})$ a. e. If $h \in \mathfrak{H}'$, $h(e^{i\theta}) \neq 0$ a. e. (See [3].) Also $\int_{|\lambda|=1} |h(\lambda)|\, |d\lambda| < \infty$ and $H(z) = \frac{1}{2\pi i} \int_{|\lambda|=1} (\lambda - z)^{-1} h(\lambda)\, d\lambda$.

LEMMA 1. *If* $[\lambda^2, g] \neq \mathbf{C}$, $g_0(\lambda) \not\equiv 0$, $|\lambda| = 1$, *then* $g_e(\lambda) = E(\lambda^2)$, E *in* $\mathfrak{A}$.

Proof of Lemma 1. Let σ be a measure on $|\lambda| = 1$ with $\int_{|\lambda|=1} \lambda^{2n} d\sigma(\lambda) = 0$, $n \geqq 0$. Since $\sigma_0(S) = -\sigma_0(-S)$, $\int_{|\lambda|=1} \lambda^{2n} d\sigma_0(\lambda) \equiv 0$, and so $\int_{|\lambda|=1} \lambda^{2n} d\sigma_e(\lambda) = 0$, $n \geqq 0$. Since $\sigma_e(S) = \sigma_e(-S)$, $\int_{|\lambda|=1} \lambda^{2k+1} d\sigma_e(\lambda) \equiv 0$. Thus

$$\int_{|\lambda|=1} \lambda^n d\sigma_e(\lambda) = 0, \quad n \geqq 0.$$

By a known theorem, [2], this implies that σ_e is absolutely continuous, $d\sigma_e(\lambda) = h(\lambda)\, d\lambda$, and $h(\lambda) \in \mathfrak{H}'$.

By hypothesis now $[\lambda^2, g] \neq \mathbf{C}$. A well-known property of the space $\mathbf{C}$ then yields a measure μ with $\mu \not\equiv 0$ and

$$\int_{|\lambda|=1} \lambda^{2n} g^m(\lambda)\, d\mu(\lambda) = 0, \qquad n, m \geqq 0. \tag{1}$$

Setting $m = 0$, we conclude from the preceding that $d\mu_e(\lambda) = a(\lambda)\, d\lambda$, $a(\lambda)$ in $\mathfrak{H}'$. Then $a(\lambda) = -a(-\lambda)$.

But the measure $d\nu(\lambda) = g(\lambda)\, d\mu(\lambda)$ also annihilates all λ^{2n}, $n \geqq 0$, as is given by (1) for $m = 1$. Hence $d\nu_e(\lambda) = m(\lambda)\, d\lambda$, $m(\lambda)$ in $\mathfrak{H}'$. Now $d\nu_e(\lambda) = g_0(\lambda)\, d\mu_0(\lambda) + g_e(\lambda)\, d\mu_e(\lambda) = g_0(\lambda)\, d\mu_0(\lambda) + g_e(\lambda) a(\lambda)\, d\lambda$. Hence $d\mu_0(\lambda) = \{m(\lambda) - g_e(\lambda) a(\lambda)\}(g_0(\lambda))^{-1} d\lambda = b(\lambda) d\lambda$. Now $g_0(\lambda) \neq 0$, all $|\lambda| = 1$, by hypothesis. Hence $b(\lambda)$ is summable on $|\lambda| = 1$. Also $b(\lambda) = b(-\lambda)$. Thus $d\mu(\lambda) = (a(\lambda) + b(\lambda))\, d\lambda = f(\lambda)\, d\lambda$, f summable, $\not\equiv 0$. We can then rewrite (1):

$$\int_{|\lambda|=1} \lambda^{2n} g^m(\lambda) f(\lambda)\, d\lambda = 0, \qquad n, m \geqq 0. \tag{1'}$$

The above considerations, applied to the measures $g^m(\lambda)f(\lambda)d\lambda$, $m = 0, 1, 2, \cdots$, yield at once: $(g^m(\lambda)f(\lambda)d\lambda)_e = \gamma_m(\lambda)d\lambda$, γ_m in $\mathfrak{H}'$. Hence

$$(g^m(\lambda)f(\lambda))_0 = \gamma_m(\lambda) \qquad \text{a. e. on } |\lambda| = 1. \tag{2}$$

An elementary computation then gives for n odd and setting $\Delta(\lambda) = \gamma_1^2 - \gamma_0\gamma_2 = g_0^2 f(\lambda)f(-\lambda)$:

$$2^n g_e^n \Delta = \sum_{\nu=0}^{(n-1)/2} C_\nu^n (2\gamma_{n+1-\nu}\gamma_{\nu+1} - \gamma_{n-\nu}\gamma_{\nu+2} - \gamma_{n+2-\nu}\gamma_\nu) = \Delta_n,$$

a. e. on $|\lambda| = 1$. Thus $g_e(\lambda) = \frac{1}{2}\Delta_1/\Delta(\lambda) = \Phi(\lambda)$ a. e. on $|\lambda| = 1$. Here Φ is meromorphic in $|\lambda| < 1$. It must be shown, of course, that $\Delta \not\equiv 0$, and we shall do this below. Now $2^n\Phi^n\Delta(\lambda) = \Delta_n(\lambda)$ a. e. on $|\lambda| = 1$. On both sides of the equation are functions analytic in $|\lambda| < 1$ and with non-tangential boundary values existing a. e. on $|\lambda| = 1$. By a theorem of Privaloff, [3], we hence get $2^n\Phi^n(\lambda)\Delta(\lambda) = \Delta_n(\lambda)$, $|\lambda| < 1$. Since Δ is a fixed function and n arbitrary, Φ can have no poles. Further,

$$|\gamma_j(z)| = \left|\frac{1}{2\pi i}\int_{|\lambda|=1} (\lambda - z)^{-1}\gamma_j(\lambda)d\lambda\right| \leqq (1 - |z|)^{-1}\int_{|\lambda|=1} |\gamma_j(\lambda)|\,|d\lambda|.$$

Since $\gamma_j(\lambda) = (g^j f)_0(\lambda)$, we get $|\gamma_j(z)| \leqq K(1 - |z|)^{-1}\|g\|^j$ where K is a constant. Hence

$$2^n|\Phi^n(z)|\,|\Delta(z)| = |\Delta_n(z)| \leqq \sum_{\nu=0}^{(n-1)/2} C_\nu^n \cdot 4K^2(1 - |z|)^{-2}\|g\|^{n+2}$$
$$\leqq K'(1 - |z|)^{-2}\cdot 2^n\|g\|^{n+2}.$$

Taking n-th roots and letting $n \to \infty$, we have $|\Phi(z)| \leqq \|g\|$, provided $\Delta(z) \neq 0$. But the zeros of Δ are isolated in $|z| < 1$ and so Φ is bounded in $|z| < 1$. Since $g_e(\lambda) = \lim_{r\to 1}\Phi(r\lambda)$ for a. a. λ, $|\lambda| = 1$, and since g_e is continuous, we conclude that $\Phi(\lambda)$ is continuous in $|\lambda| \leqq 1$. Thus $g_e(\lambda) = E(\lambda^2)$, E analytic in $|\lambda| < 1$, continuous in $|\lambda| \leqq 1$, i. e. E is in $\mathfrak{A}$.

It remains to show that $\Delta \not\equiv 0$. Suppose the contrary. Since $\Delta(\lambda) = g_0^2(\lambda)f(\lambda)f(-\lambda)$ and $g_0^2 \neq 0$, we get $f(\lambda)f(-\lambda) = 0$ a. e. on $|\lambda| = 1$. Let $D = \{\lambda \mid f$ is defined at λ and $-\lambda$, $f_0(\lambda) \neq 0$ and $f(\lambda)f(-\lambda) = 0\}$. Now $f_0 \not\equiv 0$, for else $f(\lambda) \equiv f(-\lambda)$ and so $f^2(\lambda) = 0$ a. e. Since f is defined a. e., $f(\lambda)f(-\lambda) = 0$ a. e. and $f_0 \in \mathfrak{H}'$ by (2), with $m = 0$, and so $f_0(\lambda) \neq 0$ a. e., D has measure 2π.

Let $S = \{\lambda$ in $D \mid f(\lambda) \neq 0\}$. Then $S \cup -S = D$ and $S \cap -S$ is empty, as is easily verified. Hence the Lebesgue measure of S, $m(S)$, $= m(-S) = \pi$.

By (2), $(g^m f)_0(\lambda) = \gamma_m(\lambda)$ a. e. on $|\lambda| = 1$, γ_m *in* $\mathfrak{H}'$. For λ in S,

$(g^m f)_0(\lambda) = \frac{1}{2} g^m(\lambda) f(\lambda) = g^m(\lambda) f_0(\lambda)$. Hence $g^m(\lambda) = \gamma_m(\lambda)/f_0(\lambda)$ a. e. on S. Set $G(\lambda) = \gamma_1(\lambda)/f_0(\lambda)$. Then G is meromorphic in $|\lambda| < 1$ and $G(\lambda) = g(\lambda)$ on a subset S_1 of S with $m(S_1) = m(S)$ and $G(\lambda) = g(-\lambda)$ on S_2, $S_2 \subset -S$, $m(S_2) = m(-S)$. Then $f_0(\lambda) G^m(\lambda) = \gamma_m(\lambda)$ a. e. on S. Since $m(S) > 0$, it follows that this relation is true for $|\lambda| < 1$ and a. e. on $|\lambda| = 1$. Hence G has no poles in $|\lambda| < 1$, and also

$$f_0(z) G^m(z) = \frac{1}{2\pi i} \int_{|\lambda|=1} (\lambda - z)^{-1} f_0(\lambda) G^m(\lambda) d\lambda.$$

Now $G(\lambda) = g(\lambda)$ a. e. on S, $G(\lambda) = g(-\lambda)$ a. e. on $-S$, and $m(S \cup -S) = 2\pi$. Hence $|G(\lambda)| \leqq \| g \|$ a. e. on $|\lambda| = 1$. Thus

$$|f_0(z)|\,|G^m(z)| \leqq (K/(1 - |z|))\| g \|^m,$$

whence $|G(z)| \leqq \| g \|$ if $f_0(z) \neq 0$. Thus G is bounded in $|z| < 1$.

Choose now λ_0 in $S_1^* \cap S_2^*$, (the asterisk meaning closure). There exists such a λ_0, since S_1^* and S_2^* are closed sets on the circle, each of measure π.

Set $g(\lambda_0) = \alpha$, $g(-\lambda_0) = \beta$. We shall prove $\alpha = \beta$. To this end set $H(z) = (G(z) - \alpha)(G(z) - \beta)$. $H(z)$ is then a bounded analytic function in $|z| < 1$. Given $\epsilon > 0$, choose $\delta > 0$, so that $|g(\lambda) - \alpha| < \epsilon$ and $|g(-\lambda) - \beta| < \epsilon$ provided $|\lambda - \lambda_0| < \delta$.

Now a. e. on $|\lambda| = 1$ either $G(\lambda) = g(\lambda)$ or $G(\lambda) = g(-\lambda)$. Hence a. e. on $|\lambda| = 1$, $|\lambda - \lambda_0| < \delta$, $|H(\lambda)| \leqq 2 \| g \| \cdot \epsilon$. It follows from the Poisson integral representation for H in $|z| < 1$ that $|H(z)| < \epsilon$ if $|z| < 1$ and $|z - \lambda_0| < \delta'$.

Now $\lambda_0 \in S_1^*$ and also $\lambda_0 \in S_2^*$. Hence the neighborhood $|z - \lambda_0| < \delta'$ contains z_1, z_2 with $|G(z_1) - \alpha| < \epsilon^{\frac{1}{2}}$, $|G(z_2) - \beta| < \epsilon^{\frac{1}{2}}$. An arc in this neighborhood which joins z_1 and z_2 then contains some z where both $|G(z) - \alpha| < \epsilon^{\frac{1}{2}}$ and $|G(z) - \beta| < \epsilon^{\frac{1}{2}}$. Hence $\alpha = \beta$. Thus $g(\lambda_0) = g(-\lambda_0)$. This contradicts $g_0(\lambda_0) \neq 0$. Hence $\Delta \not\equiv 0$. Lemma 1 is thus established.

Proof of Theorem 1. $[\lambda^2, \psi] \neq \mathbf{C}$ by assumption, and also $\psi_0(\lambda) \neq 0$ since λ^2, ψ separate points on $|\lambda| = 1$. Lemma 1 then gives that $\psi_e(\lambda) = E_1(\lambda^2)$ with E_1 in $\mathfrak{A}$. Now any function in $\mathfrak{A}$ may be uniformly approximated by polynomials in $|\lambda| \leqq 1$ and so ψ_e is approximable in the norm of $\mathbf{C}$ by polynomials in λ^2. Hence $\psi_e \in [\lambda^2, \psi]$. It follows that $\psi_0 \in [\lambda^2, \psi]$.

Set $h = \psi_0^2 + \psi$. Then $h_0 = \psi_0 \neq 0$ on $|\lambda| = 1$. Also $[\lambda^2, h] \subseteq [\lambda^2, \psi] \neq \mathbf{C}$. By Lemma 1, then, $h_e(\lambda) = \psi_0^2(\lambda) + \psi_e(\lambda) = E(\lambda^2)$, E in $\mathfrak{A}$ and so $\psi_0^2(\lambda) = F(\lambda^2)$, F in $\mathfrak{A}$. Now $\psi_0(\lambda) \neq 0$ if $|\lambda| = 1$ and so $F(z)$ has only finitely many zeros in $|z| < 1$. Thus we can write

$$F(z) = \prod_{i=1}^{n} (z - \lambda_i^2)\,(E_2(z))^2,$$

where $E_2 \in \mathfrak{A}$ and where the λ_i^2 are distinct points in $|\lambda| < 1$. Hence $\psi_0(\lambda) = \{\prod_{i=1}^{n} (\lambda^2 - \lambda_i^2)\}^{\frac{1}{2}} E_2(\lambda^2)$, where $\{\prod_{i=1}^{n} (\lambda^2 - \lambda_i^2)\}^{\frac{1}{2}}$ is one of the two single valued branches of this multiple-valued function defined on $|\lambda| = 1$. Since $\psi_0(\lambda) = -\psi_0(-\lambda)$, n must be odd. Thus $\psi = \psi_e + \psi_0$ has the desired representation.

Conversely, suppose ψ is of the form $\psi(\lambda) = E_1(\lambda^2) + p(\lambda)E_2(\lambda^2)$, where $E_1, E_2 \in \mathfrak{A}$, $p(\lambda) = \{\prod_{i=1}^{n} (\lambda^2 - \lambda_i^2)\}^{\frac{1}{2}}$. Then $[\lambda^2, \psi] \subseteq [\lambda^2, p]$. Now it is easily seen, using the fact that $p(\lambda) = -p(-\lambda)$, that $[\lambda^2, p]$ consists of those and only those functions having the form: $f(\lambda) = A(\lambda^2) + p(\lambda)B(\lambda^2)$, A, B in $\mathfrak{A}$. Hence clearly $[\lambda^2, p] \neq \mathbf{C}$ and so $[\lambda^2, \psi] \neq \mathbf{C}$. Theorem 1 is thus proved.

Proof of Theorem 2. Suppose $[\lambda^2, p] \subseteq \mathbf{M}'$, where $\mathbf{M}'$ is a proper closed subalgebra of $\mathbf{C}$. We must show $\mathbf{M}' = [\lambda^2, p]$.

Take any f in $\mathbf{M}'$. Let $h = f + rp$ where r is a constant chosen so that $h_0(\lambda) \neq 0$ on $|\lambda| = 1$. Since $[\lambda^2, h] \subseteq \mathbf{M}'$, Theorem 1 then gives

$$h(\lambda) = A(\lambda^2) + \{\prod_{j=1}^{m} (\lambda^2 - \xi_j^2)\}^{\frac{1}{2}} \cdot B(\lambda^2),\ A, B \text{ in } \mathfrak{A}.$$

Hence

$$h_0 \cdot p(\lambda) = \{\prod_{j=1}^{m} (\lambda^2 - \xi_j^2)\}^{\frac{1}{2}} \cdot \{\prod_{i=1}^{n} (\lambda^2 - \lambda_i^2)\}^{\frac{1}{2}} B(\lambda^2) \in \mathbf{M}'.$$

Set $k = h_0 p + p$. Then $k_0(\lambda) = p(\lambda) \neq 0$ and $[\lambda^2, k] \subseteq \mathbf{M}'$. Hence by Theorem 1, $k_e(\lambda) = h_0 p(\lambda) = K(\lambda^2)$, K in $\mathfrak{A}$. Hence the ξ_j and the λ_i must be equal in pairs, whence $h(\lambda) = A(\lambda^2) + p(\lambda)B(\lambda^2)$. Thus $h \in [\lambda^2, p]$ and so $f \in [\lambda^2, p]$. Thus $\mathbf{M}' = [\lambda^2, p]$.

2. Let now ϕ, ψ be any pair of functions in $\mathbf{C}$ separating points on $|\lambda| = 1$. Let Γ be the curve in R_4, the space of two complex variables, which is given parametrically by: $z_1 = \phi(\lambda)$, $z_2 = \psi(\lambda)$, $|\lambda| = 1$. Γ is then a simple closed Jordan curve. By a *piece of an analytic surface* we mean a bounded subset $\mathcal{F}$ of R_4 such that if $(z_1^0, z_2^0) \in \mathcal{F}$, there exist functions $z_1(\xi), z_2(\xi)$ analytic in a neighborhood of ξ_0 in the complex ξ-plane such that $z_1 = z_1(\xi), z_2 = z_2(\xi)$ represent $\mathcal{F}$ parametrically in a neighborhood of (z_1^0, z_2^0) with $z_1(\xi_0) = z_1^0$, $z_2(\xi_0) = z_2^0$. We now have condition:

(I) *There exists no piece of an analytic surface bounded by* Γ.

THEOREM 3. *Condition* **(I)** *is necessary in order that* $[\phi, \psi] = \mathbf{C}$.

Proof. Suppose $\mathfrak{F}$ is a piece of an analytic surface bounded by Γ. Let $P(z_1, z_2)$ be any polynomial. Restricted to $\mathfrak{F}$, $P(z_1, z_2)$ is an analytic function on $\mathfrak{F}$ and hence $|P(z_1, z_2)|$ attains its maximum on the boundary.

Fix now z_1^0, z_2^0 in $\mathfrak{F}$. For all polynomials P set $m(P(\phi, \psi)) = P(z_1^0, z_2^0)$. Then by the above,

$$|m(P(\phi, \psi))| \leqq \sup_{(z_1, z_2) \in \Gamma} |P(z_1, z_2)| = \|P(\phi, \psi)\|.$$

Thus m is a bounded multiplicative functional defined on a dense subset of $[\phi, \psi]$ and hence may be extended to be a multiplicative functional on all of $[\phi, \psi]$. If now $[\phi, \psi] = \mathbf{C}$, every such multiplicative functional has the form: $f \to f(\lambda_0)$, λ_0 fixed, $|\lambda_0| = 1$. Thus $z_1^0 = m(\phi) = \phi(\lambda_0)$, $z_2^0 = m(\psi) = \psi(\lambda_0)$. But $(z_1^0, z_2^0) \notin \Gamma$ and $(\phi(\lambda_0), \psi(\lambda_0)) \in \Gamma$. This is a contradiction, and so our assertion holds.

THEOREM 4. *Let* $\phi(\lambda)$ *be one-one on* $|\lambda| = 1$. *Then* **(I)** *is also sufficient in order that* $[\phi, \psi] = \mathbf{C}$.

Proof. Let γ be the curve on which $\phi(\lambda)$ maps the unit circle. Then γ is a simple closed curve in the plane. Suppose now $[\phi, \psi] \neq \mathbf{C}$. The author showed in [1] that then ψ belongs to the algebra generated by ϕ and hence that $\psi(\lambda) = F(\phi(\lambda))$, where F is continuous inside and on γ and analytic inside γ. Then the set of points (z_1, z_2) in R_4 with $z_1 = \xi$, $z_2 = F(\xi)$ where ξ ranges over the interior of γ is a piece of an analytic surface $\mathfrak{F}$ and $\mathfrak{F}$ is bounded by Γ since $(\xi, F(\xi))$ with ξ on γ is the general point on Γ.

THEOREM 5. *Let* λ^2, ψ *separate points on* $|\lambda| = 1$. *Then* **(I)** *is also sufficient in order that* $[\lambda^2, \psi] = \mathbf{C}$.

Proof. Suppose $[\lambda^2, \psi] \neq \mathbf{C}$. By Theorem 1,

$$\psi(\lambda) = E_1(\lambda^2) + \{\prod_{i=1}^{n} (\lambda^2 - \lambda_i^2)\}^{\frac{1}{2}} E_2(\lambda^2), \quad E_1, E_2 \in \mathfrak{A}, \; n \text{ odd}.$$

Let $\mathfrak{S}$ be the Riemann surface of the function $\{\prod_{i=1}^{n} (z - \lambda_i^2)\}^{\frac{1}{2}}$, represented as a two-sheeted covering of the z-plane with branch-points over λ_i^2, $i = 1, 2, \cdots, n$. Let $\mathfrak{S}'$ be the region of $\mathfrak{S}$ lying over the region $|z| < 1$, and γ its boundary. Then γ is a simple closed Jordan curve on $\mathfrak{S}$. Let Z be the function on $\mathfrak{S}'$ whose value at q, where q lies over z, $= z$, and W the function

on $\mathcal{S}'$ whose value at q is $E_1(z) + \{\prod_{i=1}^{n}(z-\lambda_i^2)\}^{\frac{1}{2}}E_2(z)$. Z and W are analytic on $\mathcal{S}'$ and together separate points on $\mathcal{S}'$.

The set of points $(Z(q), W(q))$ in R_4, q in $\mathcal{S}'$, is a piece of an analytic surface $\mathcal{F}$. Its boundary is the set of points $(Z(q), W(q))$ with q in γ. But for q in γ, q lying over λ^2, $Z(q) = \lambda^2$, $W(q) = \psi(\lambda)$, and so $(Z(q), W(q)) \in \Gamma$. Conversely all points in Γ are obtainable in this way. Hence Γ is the boundary of $\mathcal{F}$. Thus **(I)** fails if $[\lambda^2, \psi] \neq \mathbf{C}$. q. e. d.

Theorems 4 and 5 suggest the conjecture that $[\phi, \psi] = \mathbf{C}$ if and only if there exists no analytic surface bounded by Γ.

We hope to return to this question at some later date.

YALE UNIVERSITY.

REFERENCES.

[1] J. Wermer, "On algebras of continuous functions," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 866-869.

[2] C. Esseen, "Fourier analysis of distribution functions," *Acta Mathematica*, vols. 77-78 (1945-46), pp. 3-125.

[3] N. N. Lusin and I. I. Privaloff, "Sur l'unicité et multiplicité des fonctions analytiques," Annales de l'École Normale Superieur, ser. 3, vol. 42 (1925), pp. 143-191.

ON MONOTONE SOLUTIONS OF SYSTEMS OF NON-LINEAR DIFFERENTIAL EQUATIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

A theorem of Kneser [3], pp. 183-191, on a non-linear differential equation of second order,

$$y'' = F(t, y), \tag{1}$$

implies that if $q = q(t)$ is a non-negative, continuous function on $0 \leqq t < \infty$, then the linear equation

$$y'' = q(t)y \tag{2}$$

possesses a solution $y = y(t)$ satisfying $y(0) > 0$,

$$y(t) \geqq 0, \qquad y'(t) \leqq 0, \qquad y''(t) \geqq 0 \tag{3}$$

on $0 \leqq t < \infty$.

Kneser's theorem dealing with (1) has been generalized in [1] so as to apply to an equation of the type

$$y'' = F(t, y, y'); \tag{4}$$

cf. [1], where references are given to earlier treatments of (4). The theorem concerning (2) has been generalized in [2] to the case in which (2) is replaced by a linear system of differential equations of first order,

$$x' = -A(t)x, \tag{5}$$

where x is a vector with n components $x^1, \cdots, x^n$ and $A(t) = \| a_{jk}(t) \|$ is a continuous n by n matrix.

The proof in [2] for the general theorem on (5) is considerably simpler than older proofs for the particular case dealing with (2). The object of the present note is to carry over the ideas of that proof to a non-linear system

$$x' = -f(t, x), \tag{6}$$

where x and f are vectors; so that (4) is a particular case of (6).

The following terminology will be used: A vector x is called non-negative

---

* Received June 24, 1954.

($x \geqq 0$) or positive ($x > 0$) if all of its components are non-negative or positive. Similar definitions apply to $x \leqq 0$, $x < 0$, and to $x_1 \geqq x_2$ if x_1, x_2 are two vectors.

(*) *Let* $f = (f^1, \cdots, f^n) = f(t, x)$ *be a continuous (vector) function of* $(t, x) = (t, x^1, \cdots, x^n)$ *for*

$$t \geqq 0, \qquad x \geqq 0, \tag{7}$$

satisfying

$$f(t, 0) \equiv 0; \tag{8_1}$$

$$f(t, x) \geqq 0. \tag{8_2}$$

For every constant $T > 0$, *let there exist a positive, continuous function* $\phi = \phi_T(r)$, $0 \leqq r < \infty$, *satisfying*

$$\int^{\infty} dr/\phi(r) = \infty \tag{9}$$

and

$$(0 \leqq) \quad x \cdot f(t, x) \leqq \phi(|x|^2) \quad \text{if} \quad 0 \leqq t \leqq T. \tag{10}$$

Then, for any constant $c > 0$, *there exists at least one (vector) function* $x = x(t)$ *which satisfies*

$$|x(0)| = c \tag{11}$$

and which is a solution of (6) *for* $0 \leqq t < \infty$ *(so that*

$$x(t) \geqq 0 \quad \text{and} \quad x'(t) \leqq 0 \tag{12}$$

for $0 \leqq t < \infty$*).*

In (10) and (11), the symbol $|x|$ denotes the Euclidean length of the vector x and the dot denotes scalar multiplication.

(*) contains the result of [2] concerning (5); in fact, if $f(t, x) = A(t)x$, then $\phi_T(r)$ in (10) can be chosen to be $C_T r$, where $C_T = \text{const.}$, and (8_2) holds on the set (7) if the elements $a_{jk}(t)$ of $A(t)$ are non-negative.

If $x = (x^1, x^2)$ is identified with $(y, -y')$, then (4) becomes the system

$$x^{1\prime} = -x^2, \qquad x^{2\prime} = -F(t, x^1, -x^2).$$

If this system is identified with (6), so that $f = (f^1, f^2)$ becomes $(1, F(t, x^1, -x^2))$, then (*) requires that $F(t, y, y')$ be defined and continuous for $t \geqq 0$, $y \geqq 0$, $y' \leqq 0$; that

$$F(t, 0, 0) \equiv 0 \quad \text{and} \quad F(t, y, y') \geqq 0; \tag{13}$$

finally that, for every T, there should exist a positive continuous function $\phi = \phi_T(r)$ satisfying (9) and

$$(14)\quad (0\leqq)\ -yy'-y'F(t,y,y')\leqq\phi(y^2+y'^2)\ \text{ if }\ 0\leqq t\leqq T,\ y\geqq 0,\ y'\leqq 0.$$

Under these conditions (*) implies that, for every $c>0$, there is a solution of (4) satisfying $y^2(0)+y'^2(0)=c$ and (3) for $0\leqq t<\infty$.

This assertion is in some respect similar to that of (I) in [1] which, however, places more emphasis on the unequal standing of the components of the vector $x=(y,-y')$. In [1], the function $F(t,y,y')$ is continuous for $t\geqq 0$, $y\geqq 0$, $y'\leqq 0$ and satisfies (13), but (9) and (14) are replaced by the requirement that, for every $T>0$, there exists a function $\psi=\psi_T(r)$, $0\leqq r<\infty$, satisfying

$$(15)\qquad \int^{\infty} r\,dr/\psi(r)=\infty$$

and

$$(16)\qquad F(t,y,y')\leqq\psi(-y')\ \text{ if }\ 0\leqq t\leqq T,\ 0\leqq y\leqq T,\ y'\leqq 0.$$

Correspondingly, (I) in [1] permits the assignment of $y(0)>0$, rather than that of $y^2(0)+y'^2(0)$. (Note that if y is fixed (and $y'\neq 0$), then (14) implies the inequality in (16) with $\psi(-y')=-\phi(y^2+y'^2)/y'$. With this choice of ψ, the relations (9) and (15) are equivalent.)

Proof of (*). Let the definition of $f(t,x)$ be extended over the $(n+1)$-dimensional half-space ($0\leqq t<\infty$, x arbitrary) by the assignment $f(t,x)=0$ if at least one component of x is negative. It follows from (8_1) that $f(t,x)$ is continuous on this half-space. Furthermore, (8_2) and (10) hold for arbitrary x (whether or not $x\geqq 0$).

The condition (9) and the inequality (10) (or just $|x\cdot f(t,x)|\leqq\phi(|x|^2)$) imply that if $x=x(t)$ is a solution of (6) on some interval $(0\leqq)\ t_0\leqq t\leqq t^0$, then all continuations of $x(t)$ exist on $0\leqq t<\infty$. This follows from [4], pp. 176-178, if (10) is replaced by $|f(t,x)|\leqq\phi(|x|)$. In order to use (10), let $r=|x|$ in [4], p. 177, be replaced by $r=|x|^2$; cf. [5], p. 557. Then, by (10), $|r'|=|2x\cdot x'|\leqq 2|x\cdot f(t,x)|\leqq 2\phi(r)$ and so the argument applied in [4], p. 177, becomes applicable.

Note that the formulation above in terms of "all continuations" is needed since the conditions imposed on (6) do not imply the local uniqueness of solutions of (6). The assertion concerning the continuability of $x=x(t)$ over the range $0\leqq t<\infty$ is not true if $f(t,x)$ is defined only on the set (7).

Let $T>0$ and let $x_0>0$ be a given positive vector. Consider a solution $x=x(t)=x(t,T,x_0)$ of (6) determined by the initial condition $x(T)=x_0$. This solution can be considered to exist for $0\leqq t<\infty$ and, in particular, for $0\leqq t\leqq T$. In view of (6) and (8_2), $x'(t)\leqq 0$. Consequently,

$x(t) \geqq x(T) > 0$ if $0 \leqq t \leqq T$. (In particular, $x = x(t)$ on $0 \leqq t \leqq T$ is a solution of (6) even before the extension of the definition of f.)

For a given $T > 0$ and a given $c > 0$, there exist at least one $x_0 = x_0(T, c)$ and a corresponding solution $x = x(t; T) = x(t, T, x_0)$ of (6) satisfying

$$x(t; T) \geqq 0 \quad \text{on} \quad 0 \leqq t \leqq T \tag{17}$$

and, for $t = 0$,

$$| x(0; T) | = c. \tag{18}$$

If this fact is granted for a moment, the proof of (*) can be completed as follows:

Since $x = x(t; T)$ satisfies (6), hence $x'(t; T) \leqq 0$ by (8_2), it is seen from (17) and (18) that

$$| x(t; T) | \leqq c \quad \text{on} \quad 0 \leqq t \leqq T. \tag{19}$$

In view of (19) and the fact that $x = x(t; T)$ is a solution of (6), the sequence of functions $x = x(t; m)$ of t, where $m = 1, 2, \cdots$, is uniformly bounded and equicontinuous on any finite t-interval. Hence there exists a subsequence which is uniformly convergent on every finite t-interval. Let $x = x(t)$ denote the limit of such a subsequence.

By (18), $x = x(t)$ satisfies (11) at $t = 0$. Also, $x = x(t)$ is a solution of (6), since each $x = x(t; m)$ is, and satisfies $x(t) \geqq 0$ on $0 \leqq t < \infty$, since $x(t; m) \geqq 0$ on $0 \leqq t \leqq T$ if $m \geqq T$; cf. (17). Finally, $x'(t) \leqq 0$, by (6) and (8_2).

Thus, in order to complete the proof of (*), all that remains to be proved is the assertion, granted above, concerning (17) and (18). To this end, suppose first that the solutions of (6) are uniquely determined by initial conditions. Thus $x = x(t, T, x_0)$ is uniquely determined by (T, x_0) and is a continuous function of (t, T, x_0). In particular, $| x(0, T, x_0) |$ is a continuous function of x_0 for fixed T. Since (6) and (8_1) show that $x(t, T, 0) \equiv 0$, it is seen that $| x(0, T, 0) | = 0$. Also, $x_0 > 0$ implies $x(0, T, x_0) \geqq x_0$, hence $| x(0, T, x_0) | \geqq | x_0 |$. Thus, if c (> 0) is fixed, continuity considerations supply the existence of an $x_0 > 0$ satisfying $| x(0, T, x_0) | = c$. This proves the statement concerning (17) and (18) when the solutions of (6) are uniquely determined by initial conditions.

If (6) does not have this (local) property, let T be fixed and, on the bounded (t, x)-set defined by

$$0 \leqq t \leqq T; \qquad x \geqq 0, \qquad | x | \leqq 2c, \tag{20}$$

let $f_1(t,x), f_2(t,x), \cdots$ be a sequence of smooth functions satisfying

$$(21_1) \qquad f_j(t,0) \equiv 0; \qquad (21_2) \qquad f_j(t,x) \geqq 0$$

and, as $j \to \infty$,

$$(22) \qquad f_j(t,x) \to f(t,x)$$

uniformly on (20). (The existence of such functions f_j is clear, for example, from the theory of Fourier series.)

The uniformity of (22) and the fact that $\phi = \phi_T(r)$ is a positive function show that, on (20),

$$(23) \qquad (0 \leqq) \quad x \cdot f_j(t,x) < 2\phi_T(|x|^2)$$

if j is sufficiently large. Consequently, if j is sufficiently large, the function $f_j(t,x)$ can be extended to the set $(0 \leqq t \leqq T, x \geqq 0)$ so as to be continuous and satisfy (21_2) and (23).

The considerations above, dealing with (6), show that, if j is sufficiently large, then

$$(24) \qquad x' = -f_j(t,x)$$

has a solution $x = x_j(t)$ satisfying

$$(25) \qquad |x_j(0)| = c \quad \text{and} \quad x_j(t) \geqq 0, x_j'(t) \leqq 0 \quad \text{on} \quad 0 \leqq t \leqq T.$$

In particular, $|x_j(t)| \leqq c$ on $0 \leqq t \leqq T$.

The sequence of functions $x_1(t), x_2(t), \cdots$ is uniformly bounded and equicontinuous on $0 \leqq t \leqq T$. Consequently, there exists a subsequence which is uniformly convergent on $0 \leqq t \leqq T$. Let $x = x(t;T)$ denote the limit function of such a subsequence. It follows from the uniformity of the limit relation (22) on (20), that $x(t) = x(t;T)$ is a solution of (6). By (25), $x(t;T)$ satisfies (17) and (18). This completes the proof of (*).

(*) and the procedure applied in [6] supply existence theorems for solutions of (6) which are representable, for $0 \leqq t < \infty$, as Laplace integrals

$$(26) \qquad x(t) = \int_0^\infty e^{-ts} d\alpha(s), \quad \text{where} \quad d\alpha(s) \geqq 0$$

(the last inequality means that $\alpha = \alpha(t)$ is a vector function the components of which are non-decreasing on $0 \leqq t < \infty$).

(**) *Let $f = f(t,x)$ in (6) satisfy the conditions of (*) and, in addition, possess continuous partial derivatives of every order satisfying*

$$(27) \qquad (-1)^{h_0} D_t^{h_0} D_{x_1}^{h_1} \cdots D_{x_n}^{h_n} f \geqq 0 \qquad (h_k = 0, 1, \cdots)$$

on (7), *where* $D_z^h = \partial^h/\partial z^h$. *Then any solution of* (6) *satisfying* (12) *on* $0 \leqq t < \infty$ *is representable in the form* (26).

Proof. The theorem of Hausdorff-Bernstein states that the existence of an $a = a(t)$ satisfying (26) is equivalent to

$$(28_h) \qquad (-1)^h d^h x/dt^h \geqq 0 \quad \text{for} \quad 0 \leqq t < \infty,$$

where $h = 0, 1, \cdots$. The relations (12) are equivalent to (28_0) and (28_1). Differentiation of (6) gives

$$x''(t) = -\partial f/\partial t - \sum_{k=1}^{n} x^{k\prime} \partial f/\partial x^k.$$

Thus it is seen from (27) and (28_1) that (28_2) holds. A simple induction proves (28_h) for $h = 0, 1, \cdots$ and, therefore, (**).

The well known "enveloping" property of the partial sums of the series

$$e^{-t} = \sum_{m=0}^{\infty} (-t)^m/m!,$$

where $x = e^{-t}$ is the solution $x' = -x$, is the simplest illustration of the following theorem:

(***) *Let* $f(t, x)$ *in* (6) *satisfy the conditions of* (*) *and, in addition, let* $f(t, x)$ *be defined and continuous on the half-space* $(0 \leqq t < \infty, |x| < \infty)$ *and be a non-decreasing function of* x,

$$(29) \qquad f(t, x_1) \leqq f(t, x_2) \quad \text{if} \quad x_1 \leqq x_2.$$

Let $x = x(t)$ *be a solution of* (6) *satisfying* (12) *for* $0 \leqq t < \infty$. *Then the successive approximations to* $x = x(t)$, *defined by*

$$(30) \qquad x_0(t) \equiv x(0); \qquad x_m(t) = x(0) - \int_0^t f(s, x_{m-1}(s))\, ds,$$

where $m = 1, 2, \cdots$, *satisfy*

$$(31_m) \qquad (-1)^m (x_m(t) - x(t)) \geqq 0 \text{ for } 0 \leqq t < \infty,$$

where $m = 0, 1, \cdots$.

Although $f(t, x)$ is defined and continuous on the half-space $(0 \leqq t < \infty$, and x arbitrary) and (29) is required on this half-space, the inequalities (8_2) and (10) are assumed only on (7). Accordingly, it is neither stated nor assumed that the successive approximations satisfy $x_m(t) \geqq 0$. Neither is it stated or assumed that the successive approximations converge.

The conditions of (*) and the additional condition (29) of (***) are satisfied if (6) is the linear system (5) and every element $a_{jk}(t)$ of the matrix $A(t)$ is continuous and non-negative.

Proof. It is clear that

$$x_m(t) - x(t) = \int_0^t \{f(s, x(s)) - f(s, x_{m-1}(s))\}ds.$$

Hence (31_m) holds, by virtue of (29), if (31_{m-1}) holds. But $x'(t) \leqq 0$ implies that $x(t) \leqq x(0)$ for $0 \leqq t < \infty$, that is, that (31_0) holds. Thus (***) follows.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "On the non-increasing solutions of $y'' = f(x, y, y')$," *American Journal of Mathematics*, vol. 73 (1951), pp. 390-404.

[2] ——— and A. Wintner, "Linear differential and difference equations with monotone solutions," *ibid.*, vol. 75 (1953), pp. 731-743.

[3] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[4] A. Wintner, "The infinities in the non-local existence problem of ordinary differential equations," *American Journal of Mathematics*, vol. 68 (1946), pp. 173-178.

[5] ———, "Asymptotic integration constants," *ibid.*, vol. 68 (1946), pp. 553-559.

[6] ———, "On the Laplace-Fourier transcendents occurring in mathematical physics," *ibid.*, vol. 69 (1947), pp. 87-98.

THE SPECTRA OF TOEPLITZ'S MATRICES.*

By PHILIP HARTMAN and AUREL WINTNER.

---

1. Introduction. The function classes L^p will refer below to functions of a real variable ϕ on the interval $-\pi < \phi < \pi$. Two functions which differ only on a zero set will be considered to be identical.

If $f = (\cdots, f_{-1}, f_0, f_1, \cdots)$ is a vector with complex components and a finite length,

$$|f| = (\sum_{n=-\infty}^{\infty} |f_n|^2)^{\frac{1}{2}} < \infty, \tag{1}$$

let $f(\phi)$ denote the function of class L^2 having the Fourier series

$$f(\phi) \sim \sum_{n=-\infty}^{\infty} f_n e^{in\phi}. \tag{2}$$

Let $f = (f^-, f^+)$, where f^+ and f^- are the respective sequences $f^+ = (f_0, f_1, \cdots)$, $f^- = (\cdots, f_{-2}, f_{-1})$, and let the corresponding functions $f^+(\phi), f^-(\phi)$ of class L^2 be called the *interior* and *exterior* functions of $f(\phi)$,

$$f^+(\phi) \sim \sum_{n=0}^{\infty} f_n e^{in\phi} \quad \text{and} \quad f^-(\phi) \sim \sum_{n=1}^{\infty} f_{-n} e^{-in\phi}. \tag{3}$$

Finally, let $\mathfrak{L}$ and $\mathfrak{T}$ denote the Laurent and Toeplitz matrices determined by f, that is, let $\mathfrak{L} = (f_{n-m})$, where $n, m = 0, \pm 1, \cdots$, and $\mathfrak{T} = (f_{n-m})$, where $n, m = 0, 1, \cdots$.

Functions of class L^1 having Fourier series of the form (3) will also be referred to as interior and exterior functions.

If $x = (x^-, x^+)$ and u^+ have finite lengths, then $\mathfrak{L}(x^-, x^+) = (y^-, y^+)$ and $\mathfrak{T}u^+ = v^+$ are defined (without necessarily having a finite length). The relationship $\mathfrak{L}(0, x^+) = (y^-, y^+)$ implies that $\mathfrak{T}x^+ = y^+$. In addition, $(y^-, y^+) = \mathfrak{L}(x^-, x^+)$ means that $f(\phi)x(\phi) = y(\phi)$, where $y(\phi)$ is a function of class L^1 with the sequence of Fourier coefficients (y^-, y^+). In particular, if $y^+ = \mathfrak{T}x^+$, then $y^+(\phi)$ is the interior function of $f(\phi)x^+(\phi)$,

$$f(\phi)x^+(\phi) = y^+(\phi) + y^-(\phi).$$

Let $f(\phi)$ be real-valued, so that

---

* Received April 15, 1954.

$$f_n = \bar{f}_{-n}. \tag{5}$$

Correspondingly, $\mathfrak{L}$ and $\mathfrak{T}$ are Hermitian matrices. $\mathfrak{L}$ can be considered as a self-adjoint operator on the space L^2. A function $x(\phi)$ of class L^2 is in the domain $D(\mathfrak{L})$ of $\mathfrak{L}$ if $y(\phi) = f(\phi)x(\phi)$ is of class L^2; in which case $y(\phi) = \mathfrak{L}x(\phi)$. According to Toeplitz ([10]; cf. [6], pp. 152-155), λ is in the spectrum, $S(\mathfrak{L})$, of $\mathfrak{L}$ if and only if the set of ϕ-values $\{\phi | \lambda - \epsilon < f(\phi) < \lambda + \epsilon\}$ has a positive measure for every $\epsilon > 0$, while λ is in the point spectrum, $P(\mathfrak{L})$, of $\mathfrak{L}$ if and only if $\text{meas}\{\phi | f(\phi) = \lambda\} > 0$. In particular, $\mathfrak{L}$ is bounded if and only if

$$f(\phi) \text{ is bounded} \tag{6}$$

(that is, if and only if $|f(\phi)| \leqq \text{const.}$ for almost all ϕ).

According to Toeplitz (cf., e. g., [3], p. 360), when (5) is assumed, $\mathfrak{T}$ is bounded if and only if $\mathfrak{L}$ is, that is, if and only if (6) holds. In this case, $\mathfrak{T}$ can be considered as a bounded self-adjoint operator on the Hilbert space of the interior functions $x^+(\phi)$ of class L^2, where $y^+(\phi) = \mathfrak{T}x^+(\phi)$ is defined by (4). Furthermore, the least value, m, and greatest value, M, of the spectrum $S(\mathfrak{T})$ of $\mathfrak{T}$ coincide with the corresponding values belonging to $S(\mathfrak{L})$.

Because of the problem of the possibility of associating a self-adjoint operator with the matrix $\mathfrak{T}$ when (6) does not hold, the case of a bounded $\mathfrak{T}$ will be considered first (Part I); the discussion of an unbounded $\mathfrak{T}$ will be deferred to Part IV.

Part I. Bounded $\mathfrak{T}$-matrices.

2. The spectrum $S(\mathfrak{T})$. The following theorem, conjectured in [3] but verified there only in certain particular cases, will first be proved.

Let (1), (2), (5) *and* (6) *hold (so that $f(\phi)$ is bounded and real-valued). Let m and M be the (essential) lower and upper bounds of $f(\phi)$.*

(i) *The spectrum $S(\mathfrak{T})$ of $\mathfrak{T}$ is the closed interval $[m, M]$.*

(ii) *If $f(\phi)$ is not a constant, that is, if $m < M$, then the point spectrum $P(\mathfrak{T})$ of $\mathfrak{T}$ is empty.*

These assertions imply that, *under the assumptions* (1), (2), (5), (6) *and* $f \neq$ const., $\mathfrak{T}$ has a continuous spectrum only. This contrasts with the situation for Laurent matrices $\mathfrak{L}$ which can have a pure point spectrum (or a pure continuous spectrum or a mixture of continuous and point spectra).

The above assertions are somewhat surprising in view of Szegö's results [8] on the asymptotic distribution of the characteristic numbers $\lambda_{1k} \leqq \lambda_{2k} \leqq \cdots \leqq \lambda_{kk}$ of the k-th section $\mathfrak{T}_k = (f_{n-m})$, where $n, m = 0, 1, \cdots, k-1$, of $\mathfrak{T}$. Those results imply, for example, that, in the case of the function defined by $f(\phi) = \pm 1$ according as $-\pi < \phi < 0$ or $0 < \phi < \pi$, one has, as $k \to \infty$,

$$\{\text{number of } \lambda_{jk} < -1+\epsilon\}/k \to \tfrac{1}{2}, \qquad \{\text{number of } \lambda_{jk} > 1-\epsilon\}/k \to \tfrac{1}{2},$$

for every $\epsilon > 0$; while, according to Toeplitz, $-1 \leqq \lambda_{jk} \leqq 1$. Thus, for large k, about a half of the numbers $\lambda_{1k}, \cdots, \lambda_{kk}$ are near -1 and about a half are near $+1$. On the other hand, the spectrum of $\mathfrak{T}$ consists of the entire interval $-1 \leqq \lambda \leqq 1$; in particular, the numbers λ_{jk} are dense in $-1 \leqq \lambda \leqq 1$.

It will be seen in Section 7 below that (ii) is false if $f(\phi)$ is not bounded. In other words, there exist real-valued functions $f(\phi)$ of class L^2 for which the corresponding (Hermitian) Toeplitz matrix $\mathfrak{T}$ satisfies $\mathfrak{T}x^+ = 0$ for some vector $x^+ (\neq 0)$ of finite length. On the other hand, (i) can be formulated so as to remain valid for all (bounded or unbounded) Hermitian Toeplitz matrices; see Section 12.

(i) will be deduced from (iii) below in Section 4; (ii) will be proved in Section 3.

3. Proof of (ii). Suppose that $P(\mathfrak{T})$ is not empty, so that there is a point $\lambda = \lambda_0$ in $P(\mathfrak{T})$. If $\mathfrak{J}$ is the identity operator, then $\mathfrak{T} - \lambda_0 \mathfrak{J}$ is the Toeplitz matrix belonging to $f(\phi) - \lambda_0$. Hence, it can be supposed that $\lambda_0 = 0$ and that there exists an $x_0^+(\phi) \not\equiv 0$ of class L^2 such that $\mathfrak{T}x_0^+(\phi) = 0$. According to (4), this means that

$$f(\phi)x_0^+(\phi) = y^-(\phi) \sim \sum_{n=1}^{\infty} y_{-n}e^{-in\phi}. \tag{7}$$

Since $m < M$, it follows that $f(\phi) \not\equiv 0$. Furthermore, by a theorem of F. and M. Riesz, $x_0^+(\phi) \not\equiv 0$ implies that $x_0^+(\phi)$ vanishes at most on a set of measure 0. Hence $y^-(\phi) \not\equiv 0$ and so there is a least (positive) integer $n = N$ satisfying $y_{-n} \neq 0$. It can be supposed that $y_{-N} = 1$, so that

$$y^-(\phi) \sim e^{-iN\phi} + y_{-(N+1)}e^{-i(N+1)\phi} + \cdots.$$

If the relation (7) is multiplied by $e^{iN\phi}$, it is seen that $\mathfrak{T}x^+(\phi) = 1$ if $x^+(\phi)$ is the inner function $e^{iN\phi}x_0^+(\phi)$. Similarly, it is seen that $\mathfrak{T}x^+(\phi) = e^{i\phi}$ if $x^+(\phi)$ is the inner function $(e^{i(N+1)\phi} - y_{-(N+1)}e^{iN\phi})x_0^+(\phi)$. An induction shows that the range of $\mathfrak{T}$ contains $1, e^{i\phi}, e^{2i\phi}, \cdots$ and is therefore dense in

the Hilbert space of interior functions. But this implies that the self-adjoint operator $\mathfrak{T}$ cannot annihilate $x_0^+(\phi) \not\equiv 0$. This contradiction proves the assertion (ii).

Part II. Riemann's problem.

4. Equation (8). The problem of the existence of interior and exterior functions $x_0^+(\phi)$, $y_0^-(\phi)$ satisfying the equation

$$f(\phi)x_0^+(\phi) = 1 + y_0^-(\phi), \tag{8}$$

when $f(\phi)$ is given, goes back to Riemann; cf. [4], pp. 309-321 and [5], p. 210. The answer to this problem when $f(\phi)$ is real-valued, bounded and measurable and solutions $x_0^+(\phi)$ of class L^2 are desired is contained in the following assertion:

(iii) *Let $f(\phi)$ be a real-valued, bounded, measurable function on $(-\pi, \pi)$. Necessary for the equation* (8) *to have a solution $x_0^+(\phi)$ of class L^2 is that $f(\phi)$ have the following properties:*

$$\operatorname{sgn} f(\phi) \text{ is a constant } (1 \text{ or } -1); \tag{9}$$

$$1/f(\phi) \text{ is of class } L^1. \tag{10}$$

It will follow from (iv) that the converse of (iii) is true, that is, that (9) and (10) are sufficient in order that (8) have solution $x_0^+(\phi)$ of class L^2.

Remark 1. It will be seen below (Section 7) that (9) does not remain a necessary condition if the hypothesis that $f(\phi)$ is bounded is replaced, for example, by the assumption that $f(\phi)$ is of class L^2.

Remark 2. *Proof of* (i). The fact that (9) is a necessary condition for the solvability of (8), with an $x_0^+(\phi)$ of class L^2, implies (i).

In order to see this, let $m < \lambda < M$. If λ is not in the spectrum $S(\mathfrak{T})$ of $\mathfrak{T}$, then the equation $(\mathfrak{T} - \lambda\mathfrak{J})x^+(\phi) = 1$, that is, the equation $(f(\phi) - \lambda)x^+(\phi) = 1 + y^-(\phi)$, has a solution $x^+(\phi)$ of class L^2. But since $\operatorname{sgn}(f(\phi) - \lambda)$ is not a constant (almost everywhere), this leads to a contradiction. Hence λ is in $S(\mathfrak{T})$. Since $S(\mathfrak{T})$ is a closed λ-set, (i) follows.

(iv) *Let $f(\phi)$ be a function of class L^1. Sufficient for* (8) *to have a solution $x_0^+(\phi)$ of class L^2 is that $f(\phi)$ have the properties* (9) *and* (10) *(in which case $y_0^-(\phi)$ is of class L^2).*

It may be remarked that if (8) holds for functions $x_0^+(\phi)$, $y_0^-(\phi)$ of class L^p, $p > 0$, then

(11) $\log|f(\phi)|$ is of class L^1.

In fact, $\log|x_0^+(\phi)|$ and $\log|1+y_0^-(\phi)|$ are of class L^1 ([9], p. 233 for the case $p=2$ and [7], pp. 91-92, for the case $p>0$).

Theorem (iv) will be deduced from results of Szegö [8], pp. 176-177; its proof will only be sketched. Actually, (iv) can be deduced from a theorem of Szegö [9], p. 235, which implies that if $1/f(\phi)$ is of class L^1 (where f is not assumed to be of class L^1), then $1/f(\phi)$ has a factorization of the form $1/f(\phi)=\pm|x_0^+(\phi)|^2$, with an $x_0^+(\phi)$ of class L^2, if and only if $f(\phi)$ has the properties (9) and (11). This theorem of Szegö does not imply (iii), which concerns (8), that is, the more general factorization $x_0^+(\phi)/(1+y_0^-(\phi))$ for $1/f(\phi)$; on the other hand, it turns out that this more general factorization is necessarily of the form $1/f(\phi)=\pm|x_0^+(\phi)|^2$ if f is bounded.

5. Proof of (iii). Let $f(\phi)$ be bounded and measurable and let (8) have a solution $x_0^+(\phi)$ of class L^2; so that $y_0^-(\phi)$ is an exterior function of class L^2. In the Fourier expansion of $x_0^+(\phi)$, the 0-th (constant) term is not 0, for otherwise multiplication of (8) by $e^{-i\phi}$ gives a relation which shows that $\lambda=0$ is in the point spectrum $P(\mathfrak{T})$ of $\mathfrak{T}$ and that $e^{-i\phi}x_0^+(\phi)$ is a corresponding eigenfunction. By (ii), this means that $f(\phi)=0$ almost everywhere, which contradicts (8). Therefore, it is possible to write

(12) $$x_0^+(\phi)\sim c\sum_{n=0}^{\infty}a_ne^{in\phi},\quad \text{where } c\neq 0,\ a_0=1.$$

The function $1+y_0^-(\phi)$ is of class L^2 and has, therefore, a Fourier expansion,

(13) $$1+y_0^-(\phi)\sim\sum_{n=0}^{\infty}\gamma_ne^{-in\phi},\qquad \gamma_0=1.$$

Let $1=\beta_0,\beta_1,\cdots$ be the sequence of numbers defined by the relation

(14) $$\Big(\sum_{n=0}^{\infty}\gamma_nz^{-n}\Big)^{-1}=\sum_{n=0}^{\infty}\beta_nz^{-n}$$

for large $|z|$. Thus, formally,

(15) $$(1+y_0^-(\phi))^{-1}\sim\sum_{n=0}^{\infty}\beta_ne^{-in\phi},\qquad \beta_0=1.$$

Let $\mathfrak{X}$, $\mathfrak{Y}$, $\mathfrak{Y}_0$ denote the three Toeplitz matrices

(16) $$\mathfrak{X}=\begin{bmatrix}1&0&0&.\\ a_1&1&0&.\\ a_2&a_1&1&.\\ .&.&.&.\end{bmatrix},\quad \mathfrak{Y}=\begin{bmatrix}1&\beta_1&\beta_2&.\\ 0&1&\beta_1&.\\ 0&0&1&.\\ .&.&.&.\end{bmatrix},\quad \mathfrak{Y}_0=\begin{bmatrix}1&0&0&.\\ \gamma_1&1&0&.\\ \gamma_2&\gamma_1&1&.\\ .&.&.&.\end{bmatrix}$$

belonging to the functions $x_0^+(-\phi)/c$, $(1+y_0^-(-\phi))^{-1}$, $1+y_0^-(\phi)$, respectively.

Let (8) be multiplied by $e^{in\phi}$ and the result written as

$$f(\phi)x_0^+(\phi)e^{in\phi} = p_n(\phi) + y_n^-(\phi), \tag{17}$$

where $p_n(\phi) = e^{in\phi} + \gamma_1 e^{i(n-1)\phi} + \cdots + \gamma_n$ is an interior function which is a polynomial of degree n in $e^{i\phi}$ and $y_n^-(\phi)$ is an exterior function. Then (16) implies that, in terms of matrix-vector multiplication,

$$\mathfrak{Y}_0(1, e^{i\phi}, e^{2i\phi}, \cdots) = (1, p_1(\phi), p_2(\phi), \cdots).$$

The matrix $\mathfrak{Y}_0$ has a unique formal inverse, $\mathfrak{Y}^{tr}$ which is the transpose of $\mathfrak{Y}$,

$$(1, e^{i\phi}, e^{2i\phi}, \cdots) = \mathfrak{Y}^{tr}(1, p_1(\phi), p_2(\phi), \cdots);$$

that is,

$$e^{in\phi} = \sum_{j=0}^{n} \beta_{n-j} p_j(\phi), \qquad p_0(\phi) \equiv 1; \tag{18}$$

cf., e. g., [12], p. 280. It follows from (17) and (18) that

$$f(\phi)x_0^+(\phi) \sum_{j=0}^{n} \beta_{n-j} e^{ij\phi} = e^{in\phi} + y_{(n)}^-(\phi),$$

where $y_{(n)}^-(\phi)$ is an exterior function.

Let $x^+ = (x_0, x_1, \cdots)$ be a vector having only a finite number of components different from 0 and let the last relation be multiplied by x_n and the result summed with respect to n. There results the equation

$$f(\phi)x_0^+(\phi) \sum_{j=0}^{\infty} \Big(\sum_{n=0}^{\infty} \beta_n x_{n+j} \Big) e^{ij\phi} = x^+(\phi) + y^-(\phi), \tag{19}$$

in which the double series is a finite sum and $y^-(\phi)$ is an exterior function.

Since the case $f(\phi) \not\equiv 0$ is excluded, $\lambda = 0$ is not in the point spectrum of $\mathfrak{T}$. Hence $\mathfrak{T}^{-1}$ exists and is a self-adjoint operator, which may or may not be bounded. Equation (19) implies

$$\mathfrak{T}^{-1}x^+(\phi) = x_0^+(\phi) \sum_{j=0}^{\infty} \Big(\sum_{n=0}^{\infty} \beta_n x_{n+j} \Big) e^{ij\phi}. \tag{20}$$

It follows therefore from (12) and (14) that

$$\mathfrak{T}^{-1}x^+ = c\mathfrak{X}(\mathfrak{Y}x^+) = c(\mathfrak{Y}\mathfrak{X})x^+. \tag{21}$$

In fact, the legitimacy of the necessary rearrangement of summations is trivial, because of the triangular forms of $\mathfrak{X}$ and $\mathfrak{Y}$ and the fact that the vector x^+ has only a finite number of non-vanishing components. Since $\mathfrak{T}^{-1}$ is self-

adjoint (in particular, the scalar product relation $(\mathfrak{T}^{-1}x_1^+, x_2^+) = (x_1^+, \mathfrak{T}^{-1}x_2^+)$ holds if x_1^+ and x_2^+ are vectors with only a finite number of non-vanishing components), it follows that the matrix $c\mathfrak{X}\mathfrak{Y}$ is Hermitian. Since the first row of $\mathfrak{X}\mathfrak{Y}$ is $1, \beta_1, \cdots$ and its first column is $1, a_1, \cdots$,

$$c \text{ is real and } \beta_n = \bar{a}_n, \qquad n = 0, 1, \cdots. \tag{22}$$

Since $x_0^+(\phi)$ is of class L^2, (12) and (18) show that $(1 + y_0^-(\phi))^{-1}$ is of class L^2 and, in fact, that $(1 + y_0(\phi))^{-1}$ is the complex conjugate of $x_0^+(\phi)/c$. Hence (8) implies that $f(\phi)$ can vanish only on a zero set and that

$$1/f(\phi) = |x_0^+(\phi)|^2/c. \tag{23}$$

Thus, according as $c > 0$ or $c < 0$, $f(\phi) > 0$ or $f(\phi) < 0$ almost everywhere. Finally, $1/f(\phi)$ is of class L^1, since $x_0^+(\phi)$ is of class L^2. This proves (iii).

6. Proof of (iv). Let $f(\phi) > 0$ almost everywhere and let $-\log f(\phi)$ be of class L^1. In particular, $-\log f(\phi)$ has a Fourier expansion,

$$-\log f(\phi) = g^+(\phi) + g^-(\phi) \sim \sum_{n=0}^{\infty} g_n e^{in\phi} + \sum_{n=1}^{\infty} g_{-n} e^{-in\phi}, \tag{24}$$

where

$$g_n = \bar{g}_{-n}. \tag{25}$$

For $|z| < 1$, put

$$G(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (-\tfrac{1}{2}\log f(\phi))\,(e^{i\phi} + z)\,(e^{i\phi} - z)^{-1} d\phi, \tag{26}$$

so that $G(z) = \frac{1}{2}g_0 + \sum_{n=1}^{\infty} g_n z^n$, and put

$$H^+(z) = \exp(\tfrac{1}{2}g_0 + G(z)). \tag{27}$$

According to [8], pp. 176-177, the fact that $1/f(\phi)$ is of class L^1 implies that

$$\lim_{r \to 1} \int_{-\pi}^{\pi} |H^+(re^{i\phi})|^2 \, d\phi < \infty.$$

Consequently,

$$x_0^+(\phi) = \exp g^+(\phi) \tag{28}$$

is the interior function $H^+(e^{i\phi})$ and is of class L^2. Since $f(\phi)$ is of class L^1,

it is seen in the same way that the function $y_0^-(\phi)$ defined by $1 + y_0^-(\phi) = \exp(-g^-(\phi))$ is an exterior function of class L^2. On writing $f(\phi)$ as $\exp(-g^+(\phi) - g^-(\phi))$, it follows that (8) holds. This proves (iv).

Remark 1. The construction used in this proof goes back to Plemelj [5], p. 210. This becomes clear if it is noted that the regular function $H^+(z)$, satisfying $x_0^+(\phi) = H^+(e^{i\phi})$, is given by

$$(29^+) \qquad H^+(z) = \exp \frac{1}{2\pi i} \int_{-\pi}^{\pi} (-\log f(\phi))\,(e^{i\phi} - z)^{-1} de^{i\phi}, \qquad |z| < 1.$$

and that $1 + y_0^-(\phi) = H^-(e^{i\phi})$, where $H^-(z)$ is the regular function

$$(29^-) \qquad H^-(z) = \exp \frac{1}{2\pi i} \int_{-\pi}^{\pi} (-\log f(\phi))\,(e^{i\phi} - z)^{-1} de^{i\phi}, \qquad |z| > 1.$$

Remark 2. Let $f(\phi)$ be bounded and positive almost everywhere and, in addition, let $f(\phi)$ satisfy (10). Then (8) has a solution $x_0^+(\phi)$ of class L^2 and $1/f(\phi)$ possesses the factorization (23), where

$$c = \exp\left(-\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\phi)\, d\phi\right);$$

cf. (12) and (28). Correspondingly, the factorization (21) for $\mathfrak{T}^{-1}$ can be written as

$$(30) \qquad \mathfrak{T}^{-1} = c\mathfrak{Y}^*\mathfrak{Y},$$

where $\mathfrak{Y} = \mathfrak{X}^*$ is the complex conjugate transpose of the matrix $\mathfrak{X}$ defined by (12) and (16).

The Toeplitz matrix $\mathfrak{Y}$ is bounded if and only if the function $H^+(z)$, given by (26) and (27), is bounded for $|z| < 1$ (cf., e. g., Appendix below; for references, see [12], p. 278). This is the case if and only if the real part of $G(z)$ is bounded from above for $|z| < 1$. Since the real part of $G(z)$ is the Poisson integral of $-\frac{1}{2}\log f(\phi)$, it follows that $\mathfrak{Y}$ is bounded if and only if $-\frac{1}{2}\log f(\phi) \leqq$ const. almost everywhere, that is, if and only if $f(\phi) \geqq \text{Const.} > 0$ almost everywhere. This gives another proof of the fact that if λ is not on the interval $[m, M]$, then λ is not in the spectrum of $\mathfrak{T}$.

Part III. A counter-example.

7. A pathological case. An example proving the following assertions will be given:

Let $f(\phi)$ be real-valued and of class L^2 and let $\mathfrak{T}$ be the corresponding (Hermitian) Toeplitz matrix.

(ii*) *It is possible that equation* (7) *has a solution $x_0^+(\phi)$ of class L^2, that is, that $\lambda = 0$ is in the point spectrum of $\mathfrak{T}$.*

(iii*) *It is possible that equation* (8) *has a solution $x_0^+(\phi)$ of class L^2, even though* (9) *does not hold.*

It will be seen in the proof of (iii*) that if $f(\phi)$ is real-valued, bounded and measurable but does not have the property (9), then (8) can have a solution $x_0^+(\phi)$ of class L^p for every $p < 2$ (cf. (33)-(34) below), although, according to (iii), (8) cannot have a solution $x_0^+(\phi)$ of class L^2.

8. Proof of (iii*). Let $F(z)$ denote the regular function

$$F(z) = \{(i+z)/(i-z)\}^{\frac{1}{2}} \text{ for } |z| < 1 \qquad (F(0) = 1). \tag{31}$$

Then

$$\lim_{r \to 1} \int_{-\pi}^{\pi} |F(re^{i\phi})|^p \, d\phi < \infty \text{ for every } p < 2. \tag{32}$$

Furthermore, $F(e^{i\phi})$ is (defined and) continuous except at $\phi = \frac{1}{2}\pi$ and vanishes only at $\phi = -\frac{1}{2}\pi$. On the other hand, the real part of $(F(e^{i\phi}))^2$ is identically 0, so that $u^2(\phi) \equiv v^2(\phi)$ if $F(e^{i\phi}) = u + iv$, where $u(\phi)$, $v(\phi)$ are real-valued. Put

$$f_1(\phi) = \mp 1 \text{ according as } |\phi| < \tfrac{1}{2}\pi \text{ or } \tfrac{1}{2}\pi < |\phi| < \pi. \tag{33}$$

Then $f_1(\phi) = \operatorname{sgn} v(\phi)/u(\phi)$ and $f_1(\phi)(v - iu) = u - iv$ except at $\phi = \pm \frac{1}{2}\pi$. Hence, if $x_1(\phi)$ denotes the interior function $-iF(e^{i\phi}) = v - iu$ of class L^p, $p < 2$,

$$f_1(\phi) x_1^+(\phi) = 1 + y_1^-(\phi), \tag{34}$$

where $1 + y_1^-(\phi) = u - iv$ and $y_1^-(\phi)$ is an exterior function, since $u - iv$ is the complex conjugate of $F(e^{i\phi})$ and $F(0) = 1$.

Let $0 < \epsilon < \frac{1}{2}$ and put

$$f_2(\phi) = 1/|(i - e^{i\phi})(i + e^{i\phi})|^{\epsilon}. \tag{35}$$

Then $f_2(\phi)$ is of class L^2 and $1/f_2(\phi)$ is bounded. Hence, assertion (iv) and its proof show that there is a regular function, say $H_2^+(z)$, on $|z| < 1$ which is bounded on $|z| < 1$ and has a boundary function $x_2^+(\phi) = H_2^+(e^{i\phi})$ satisfying

$$f_2(\phi)x_2^+(\phi) = 1 + y_2^-(\phi), \tag{36}$$

where $y_2^-(\phi)$ is an exterior function, and

$$|x_2^+(\phi)|^2 = \text{const.}/f_2(\phi) = \text{const.}\,|(i - e^{i\phi})(e + e^{i\phi})|^{\epsilon}. \tag{37}$$

Put $f(\phi) = f_1(\phi)f_2(\phi)$, so that, by (33) and (35), $f(\phi)$ is of class L^2. Let $x_0^+(\phi) = x_1^+(\phi)x_2^+(\phi) = -iF(e^{i\phi})H_2^+(e^{i\phi})$. Then (35) and (36) show that (8) holds with $y_0^-(\phi) = y_1^-(\phi) + y_2^-(\phi) + y_1^-(\phi)y_2^-(\phi)$. Since $x_2^+(\phi)$ is bounded, the function $x_0^+(\phi)$ is of class L^p for every $p < 2$; in fact,

$$\lim_{r\to 1} \int_{-\pi}^{\pi} |F(re^{i\phi})H_2^+(re^{i\phi})|^p \, d\phi < \infty \quad \text{if } p < 2.$$

On the other hand, the boundary function $x_0^+(\phi)$ is of class L^2, by (31) and (37), and so

$$\lim_{r\to 1} \int_{-\pi}^{\pi} |F(re^{i\phi})H_2^+(re^{i\phi})|^2 \, d\phi < \infty.$$

Hence $x_0^+(\phi)$ is an interior function of class L^2. It is seen in a similar manner, that $y_0^-(\phi)$ is an exterior function of class L^1. In view of (33), this proves (iii*).

9. Proof of (ii*). In order to prove (ii*), it is sufficient to show that for the function $f(\phi) = f_1(\phi)f_2(\phi)$, just constructed, the equation (8) has at least two solutions $x_0^+(\phi)$ of class L^2. For, if this is the case, the difference of these solutions of (8) is a solution of (7).

In order to construct another solution of (8), note that since $f_1(\phi) = \operatorname{sgn}(v/u)$ and $u^2 = v^2$, (34) holds if $x_1^+(\phi) = (v + iu)/(u^2 + v^2)$ denotes the interior function $i/F(e^{i\phi}) = i/(u + iv)$ and $1 + y_1^-(\phi)$ the function $(u + iv)/(u^2 + v^2)$, which is the complex conjugate of $1/F(e^{i\phi})$. Thus if $x_2^+(\phi)$ is the function satisfying (36) and (37), it follows that $x_0^+(\phi) = x_1^+(\phi)x_2^+(\phi) = iH_2^+(e^{i\phi})/F(e^{i\phi})$ is a solution of (8), where $1 + y_0^-(\phi)$

$= (1 + y_1^-(\phi))(1 + y_2^-(\phi))$. The arguments used after (37) show that $x_0^+(\phi)$ is an interior function of class L^2 and that $y_0^-(\phi)$ is an exterior function of class L^1.

Consequently, (8) has both functions $x_0^+(\phi) = -iF(e^{i\phi})H_2^+(e^{i\phi})$, $x_0^+(\phi) = iH_2^+(e^{i\phi})/F(e^{i\phi})$ of class L^2 as solutions. Thus (ii*) is proved; in fact, the difference $x_0^+(\phi) = -iH_2^+(e^{i\phi})(F(e^{i\phi}) + 1/F(e^{i\phi})) \not\equiv 0$ is a solution of (7) and of class L^2.

Remark. It is of interest to note that the functions $x_0^+(\phi)$, $1 + y_0^-(\phi)$ satisfying (8) in the above proof are the boundary functions $H^+(e^{i\phi})$, $H^-(e^{i\phi})$ of the functions (29$^\pm$), if $\log f(\phi)$ is suitably interpreted. Thus, in the formula $\log f(\phi) = \log f_1(\phi) + \log f_2(\phi)$, let $\log f_2(\phi)$ be real and let $\log f_1(\phi)$ be 0 or $i\pi$ according as $f_1(\phi) = 1$ or $f_1(\phi) = -1$. The resulting function (29$^+$) has the boundary values $H^+(e^{i\phi}) = -iF(e^{i\phi})H_2^+(e^{i\phi})$, that is, the first solution of (8) constructed above. On the other hand, if $\log f_1(\phi)$ is $2\pi i$ or $i\pi$ according as $f_1(\phi) = 1$ or $f_1(\phi) = -1$, the resulting function (29$^+$) has the boundary values $H^+(e^{i\phi}) = iH_2^+(e^{i\phi})/F(e^{i\phi})$, that is, the second solution of (8) constructed above.

On the other hand, not all solutions $x_0^+(\phi)$ of class L^2 of (8) result in this manner. For example, if $x_0^+(\phi) \not\equiv 0$ is a solution of (7), then for a suitable integer $N > 0$ and a suitable const., the function const. $e^{iN\phi}x_0^+(\phi)$ is a solution of (8); cf. the proof of (ii). But const. $e^{iN\phi}x_0^+(\phi)$, where $|\phi| < \pi$, are the boundary values of a function which is regular for $|z| < 1$ and which vanishes in the N-th order at $z = 0$, while a function (29$^+$) does not vanish for $|z| < 1$ (if $\log f(\phi)$ is of class L^1).

Part IV. Unbounded $\mathfrak{T}$-matrices.

10. The Toeplitz operator $\mathfrak{T}$. With a given real-valued $f(\phi)$ of class L^2 and its Toeplitz matrix $\mathfrak{T}$, associate the following operators on the Hilbert space of interior functions $x^+(\phi)$ of class L^2:

Let $D(\mathfrak{T})$ denote the set of those $x^+(\phi)$ of class L^2 having the property that $y^+(\phi)$ in (4) is of class L^2 and let $\mathfrak{T}$ denote the operator, with domain $D(\mathfrak{T})$, defined by $\mathfrak{T}x^+(\phi) = y^+(\phi)$. (This is equivalent to saying that $D(\mathfrak{T})$ is the set of those vectors x^+ of finite length with the property that $y^+ = \mathfrak{T}x^+$ is of finite length and that $\mathfrak{T}$ is the operator, with domain $D(\mathfrak{T})$, given by $\mathfrak{T}x^+ = y^+$.)

Let $D(\mathfrak{T}_0)$ denote the set of those $x^+(\phi)$ which have only a finite number

of Fourier constants distinct from 0; equivalently, $D(\mathfrak{T}_0)$ is the set of vectors x^+ having only a finite number of components different from 0. Let $\mathfrak{T}_0$ denote the operator which is the contraction of $\mathfrak{T}$ and has $D(\mathfrak{T}_0)$ as its domain. Then $\mathfrak{T}_0$ is a symmetric operator and its adjoint is $\mathfrak{T}$,

$$\mathfrak{T}_0^* = \mathfrak{T}. \tag{38}$$

Let $\mathfrak{T}^*$ denote the adjoint of $\mathfrak{T}$ and $D(\mathfrak{T}^*)$ the domain of $\mathfrak{T}^*$. Then $\mathfrak{T}^*$ is the least closed operator containing $\mathfrak{T}_0$, and $\mathfrak{T}^*$ is a closed, symmetric operator. Finally,

$$\mathfrak{T}_0 \subset \mathfrak{T}^* \subset \mathfrak{T}; \tag{39}$$

$$\mathfrak{T}^{**} = \mathfrak{T}. \tag{40}$$

While the Laurent operator $\mathfrak{L}$ associated with a real-valued $f(\phi)$ of class L^2 is always self-adjoint, the Toeplitz operator $\mathfrak{T}$ need not be self-adjoint.

11. Self-adjoint $\mathfrak{T}$. A careful perusal of the proofs of (i), (ii) and (iii) leads to the following statement:

(I) *If the assumption that $f(\phi)$ is bounded is replaced by the assumption that $f(\phi)$ has the properties that $f(\phi)$ is of class L^2 and that $\mathfrak{T}$ is self-adjoint, then the assertions of* (i), (ii) *and* (iii) *remain valid.*

There exist unbounded functions $f(\phi)$ which satisfy the assumptions of (I). This is clear from the following assertion:

(II) *If $f(\phi)$ is real-valued, of class L^2 and half-bounded, then $\mathfrak{T}$ is self-adjoint.*

In order to verify this statement, it can be supposed, without loss of generality, that $f(\phi) \geqq \text{const.} > 0$. Then (iv), the proof of (iii) and the Remark 2 in Section 6 show that $\mathfrak{T}^{-1}$ exists, is a bounded operator and has a factorization (30), where $c > 0$ and $\mathfrak{H}$ is a bounded (triangular) Toeplitz matrix. Hence $\mathfrak{T}^{-1}$ is self-adjoint. Thus (II) follows.

Remark. The consideration of Toeplitz's matrices was initiated by Toeplitz [11] in dealing with the problem of characterizing the Fourier constants of a non-negative function $f(\phi)$. It is curious that the Toeplitz operator $\mathfrak{T}$ associated with such an $f(\phi)$ of class L^2 is always self-adjoint, by (II).

12. Non-self-adjoint $\mathfrak{T}$. The function $f(\phi) = f_1(\phi) f_2(\phi)$ constructed in Section 9 shows that a Toeplitz operator $\mathfrak{T}$ need not be self-adjoint. In contrast to (ii), the following is the situation in such a case (since $\mathfrak{T}$ is the adjoint of a symmetric operator):

Whenever $\mathfrak{T}$ is not self-adjoint, every (non-real) number λ in at least one of the half-planes $\operatorname{Im}\lambda > 0$ or $\operatorname{Im}\lambda < 0$ is an eigenvalue of $\mathfrak{T}$.

Actually, in the particular case of the function $f(\phi) = f_1(\phi)f_2(\phi)$ of Section 8, every non-real number $\lambda(\operatorname{Im}\lambda > 0$ and $\operatorname{Im}\lambda < 0)$ is an eigenvalue of $\mathfrak{T}$. For, by (33) and (35), both $f_1(\phi)$ and $f_2(\phi)$ are even functions of ϕ, and so $f(\phi)$ is an even function of ϕ. Consequently, $\mathfrak{T}$ is a real matrix, so that $(\mathfrak{T} - \lambda\mathfrak{J})x^+ = 0$ implies $(\mathfrak{T} - \bar{\lambda}\mathfrak{J})\bar{x}^+ = 0$. Since $\lambda = 0$ is in the point spectrum of $\mathfrak{T}$, the Corollary on p. 230, of [1] leads to the following:

(I*) *There exists a real-valued function $f(\phi)$ of class L^2 such that the corresponding Toeplitz operator $\mathfrak{T}$ is not self-adjoint but possesses a self-adjoint contraction having a non-empty point spectrum.*

In contrast to (ii), assertion (i) can be formulated so as to be valid for all Hermitian Toeplitz matrices $\mathfrak{T}$, whether or not $\mathfrak{T}$ is self-adjoint.

(III) *Let $f(\phi)$ be a real-valued function of class L^2 and let m $(\geqq -\infty)$ and M $(\leqq \infty)$ be its (essential) lower and upper bounds. If $m < \mu < M$, then $\mathfrak{T} - \mu\mathfrak{J}$ does not possess a bounded inverse, that is, $\lambda = 0$ is in the spectrum of the self-adjoint operator $(\mathfrak{T}^* - \mu\mathfrak{J})(\mathfrak{T} - \mu\mathfrak{J})$.*

It is clear from (I) that if either m or M is finite, then $\mu = m$ or $\mu = M$ is allowed in this theorem.

The equivalence of the two forms of the assertion of (III) is a classical result of Toeplitz. Whether or not $\lambda = 0$ is in the spectrum of $(\mathfrak{T} - \mu\mathfrak{J})(\mathfrak{T}^* - \mu\mathfrak{J})$ will remain undecided.

13. Proof of (III). In view of (I), it is sufficient to prove (III) in the case that $\mathfrak{T}$ is not self-adjoint (so that $m = -\infty$ and $M = \infty$) and that $\mu = 0$.

Suppose, if possible, that $\mathfrak{T}$ is not self-adjoint but that $\lambda = 0$ is not in the spectrum of $\mathfrak{T}^*\mathfrak{T}$. In particular, $\lambda = 0$ is not in the point spectrum of $\mathfrak{T}$.

For any closed operator $\mathfrak{T}$ possessing an adjoint $\mathfrak{T}^*$, the spectra of $\mathfrak{T}^*\mathfrak{T}$ and $\mathfrak{T}\mathfrak{T}^*$ are identical except for the possible occurrence of $\lambda = 0$ in the point spectrum of one of the operators $\mathfrak{T}^*\mathfrak{T}$, $\mathfrak{T}\mathfrak{T}^*$ but not in the point spectrum of the other; [1], p. 234. Thus, for the case at hand, where $\mathfrak{T} \supset \mathfrak{T}^*$ and $\lambda = 0$ is not in the point spectrum of $\mathfrak{T}^*\mathfrak{T}$, the spectra of $\mathfrak{T}^*\mathfrak{T}$ and $\mathfrak{T}\mathfrak{T}^*$ are identical. Hence $\lambda = 0$ is not in the spectrum of $\mathfrak{T}\mathfrak{T}^*$.

Consequently, the closed, symmetric operator $\mathfrak{T}^*$ has an inverse $(\mathfrak{T}^*)^{-1}$, which is closed and bounded on its domain. The domain of $(\mathfrak{T}^*)^{-1}$, that is, the range of $\mathfrak{T}^*$, is dense in the Hilbert space (of interior functions $x^+(\phi)$

or vectors x^+), for otherwise $\lambda = 0$ is an eigenvalue of the adjoint $(\mathfrak{T}^*)^* = \mathfrak{T}$. Since $(\mathfrak{T}^*)^{-1}$ is closed and bounded on its domain, the latter is a closed set. In other words, the range of $\mathfrak{T}^*$ is the entire Hilbert space. Consequently, $\mathfrak{T}^*$ is self-adjoint. This leads to the contradiction that $\mathfrak{T}$ is self-adjoint.

The fact that $\mathfrak{T}$ is a Toeplitz operator was not used in the proof of (III). Hence, the following theorem has also been proved:

Let $\mathfrak{T}$ be the adjoint of a closed, symmetric operator $\mathfrak{T}^*$. If $\mathfrak{T}$ is not self-adjoint, then, for every real μ, the spectrum of $(\mathfrak{T}^* - \mu\mathfrak{J})(\mathfrak{T} - \mu\mathfrak{J})$ contains $\lambda = 0$.

This theorem is known if $\mathfrak{T}^*$ possesses self-adjoint extensions (Theorem (iv) in [1], p. 236) and was first proved for certain differential operators in [2].

Appendix.

In Section 1, a matrix $\mathfrak{T}$ was called a Toeplitz matrix (or a $\mathfrak{T}$-matrix) if it has the form $\mathfrak{T} = (f_{n-m})$, where $n, m = 0, 1, 2, \cdots$. Originally, Toeplitz had only considered the Hermitian case (5), $f_{-n} = \bar{f}_n$. Later on, he also considered the case where $f_n = 0$ if $n < 0$, so that the matrix is of the triangular form

$$\mathfrak{P} = \begin{bmatrix} f_0 & f_1 & f_2 & \cdot \\ 0 & f_0 & f_1 & \cdot \\ 0 & 0 & f_0 & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}. \tag{41}$$

A matrix of this form Toeplitz called a $\mathfrak{P}$-matrix.

In what follows, it is assumed that (1) holds but it is not assumed that the corresponding $\mathfrak{T}$-matrix is Hermitian or of the form (41).

(*) A $\mathfrak{T}$-matrix $\mathfrak{T} = \mathfrak{T}(f)$ belonging to a function $f(\phi)$ of class L^2 is bounded if and only if $f(\phi)$ is bounded.

Actually, (*) is a consequence of Toeplitz's assertion on the Hermitian cases. In order to see this, note that a matrix $\mathfrak{T}$ is bounded if and only if its complex conjugate transpose $\mathfrak{T}^*$ is. Put

$$\mathfrak{T} = \mathfrak{T}_1 + i\mathfrak{T}_2, \quad \text{where} \quad \mathfrak{T}_1 = \tfrac{1}{2}(\mathfrak{T} + \mathfrak{T}^*), \quad \mathfrak{T}_2 = -\tfrac{1}{2}i(\mathfrak{T} - \mathfrak{T}^*). \tag{42}$$

Hence, it is seen that $\mathfrak{T}$ is bounded if and only if $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are. On the other hand, if $\mathfrak{T} = \mathfrak{T}(f)$, then $\mathfrak{T}_1 = \mathfrak{T}(\mathrm{Re}\, f)$ and $\mathfrak{T}_2 = \mathfrak{T}(\mathrm{Im}\, f)$ are $\mathfrak{T}$-matrices belonging to real functions and, hence, are Hermitian. Thus, according to Toeplitz, $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are bounded if and only if $\mathrm{Re}\, f(\phi)$ and $\mathrm{Im}\, f(\phi)$ are bounded, respectively. Thus (*) follows.

It was conjectured in [3], p. 361, that the spectrum $S(\mathfrak{T})$ of a bounded Hermitian matrix $\mathfrak{T}$ (belonging to a real-valued $f(\phi)$) is associated with the distribution function of the (real-valued) harmonic function

$$F(r,\phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\,(1-r^2)\,(1-2r\cos(\theta-\phi)+r^2)^{-1}d\theta, \qquad (r<1), \tag{43}$$

that is,

$$F(r,\phi) = \sum_{n=-\infty}^{\infty} f_n r^{|n|} e^{in\phi}, \qquad (r<1). \tag{44}$$

This conjecture has now been justified in a certain sense. For, according to (i), $S(\mathfrak{T})$ is the closure $[m, M]$ of the range of values of $F(r,\phi)$ on the unit circle $r<1$; cf. (I) and (III) for the case of an unbounded $f(\phi)$.

From this point of view, (i) is the analogue of a result in [12], p. 279, stating that the spectrum $S(\mathfrak{P})$ of (41) is the closure of the set of values of the corresponding function (44) on $r<1$. In the case of the $\mathfrak{P}$-matrix (41), where $f_n = 0$ if $n<0$, (44) is the power series

$$F(r,\phi) = F(z) = \sum_{n=0}^{\infty} f_n z^n, \qquad (z = re^{i\phi}). \tag{45}$$

According to [12], pp. 279-280, if the complex number μ is not in the closure of the set of values of (45) for $|z|<1$, then the $\mathfrak{P}$-matrix belonging to the sequence of coefficients $h^+ = (h_0, h_1, \cdots)$ of the power series

$$(F(z)-\mu)^{-1} = \sum_{n=0}^{\infty} h_n z^n, \qquad |z|<1,$$

is bounded and is the inverse of $\mathfrak{P}-\mu\mathfrak{J}$. On the other hand, if $F(z)=\mu$ for some z, $|z|<1$, then $\lambda=\mu$ is an eigenvalue of $\mathfrak{P}$ and $x^+ = (1, z, z^2, \cdots)$ is a corresponding eigenvector. (This last fact contrasts with (ii), according to which a bounded Hermitian $\mathfrak{T}$ matrix has no eigenvalue, unless $f=$ const.)

The problem of determining the spectrum $S(\mathfrak{T})$ of a $\mathfrak{T}$-matrix which is neither Hermitian nor triangular will remain open.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "On the essential spectra of symmetric operators," *American Journal of Mathematics*, vol. 75 (1953), pp. 229-240.

[2] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[3] ——— and A. Wintner, "On the spectra of Toeplitz's matrices," *American Journal of Mathematics*, vol. 73 (1950), pp. 359-366.

[4] D. Hilbert, "Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. III," *Göttinger Nachrichten*, 1905, pp. 307-338.

[5] J. Plemelj, "Ein Erganzungssatz zur Cauchyschen Integraldarstellung analytischer Funktionen, Randwerte betreffend," *Monatshefte für Mathematik und Physik*, vol. 19 (1908), pp. 205-210.

[6] F. Riesz, *Les systèmes d'équations linéaires à une infinité d'inconnues*, Paris, 1913.

[7] ———, "Ueber die Randwerte einer analytischen Funktion," *Mathematische Zeitschrift*, vol. 18 (1923), pp. 87-95.

[8] G. Szegö, "Beiträge zur Theorie der Toeplitzschen Formen. I," *ibid.*, vol. 6 (1920), pp. 167-202.

[9] ———, "Ueber die Randwerte einer analytichen Funktion," *Mathematische Annalen*, vol. 84 (1921), pp. 232-244.

[10] O. Toeplitz, "Zur Theorie der quadratische Formen von unendlichvielen Veränderlichen," *Göttinger Nachrichten*, 1910, pp. 489-506.

[11] ———, "Ueber die Fourier'sche Entwickelung positiver Funktionen," *Rendiconti del Circolo Matematico di Palermo*, vol. 32 (1911), pp. 191-192.

[12] A. Wintner, "Zur Theorie der beschränkten Bilinearformen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 228-284.

SOLUTION OF SOME PROBLEMS OF DIVISION.* [1]

PART I. Division by a Polynomial of Derivation.

By LEON EHRENPREIS.

---

1. Introduction. The purpose of this paper is to prove that every polynomial of derivation with constant coefficients P is completely inversible as regards distributions of finite order, i. e. for any distribution S of finite order there is a distribution T of finite order such that

$$PT = S. \tag{1.1}$$

This means that division, in the sense of convolution multiplication, by a polynomial of derivation is always possible. The particular choice $S = \delta$ (Dirac's measure) shows that every constant coefficient partial differenial equation possesses an elementary solution. We show also that if S is an indefinitely differentiable function, T can be chosen to be an indefinitely differentiable function in (1.1).

The above results are extended to systems of partial differential equations with constant coefficients.

The method of proof is the following: We study the space of Fourier transforms of the space $\mathcal{D}$ of Schwartz. Call this set $\boldsymbol{D}$; we give it the topology to make the Fourier transform a topological isomorphism. $\boldsymbol{D}$ consists of all entire functions of exponential type which are rapidly decreasing by a theorem of Paley and Weiner (see [4] and [5]). The Fourier transform then defines, in a natural manner, a topological isomorphism of $\mathcal{D}'$ (the space of distributions) onto $\boldsymbol{D}'$ (the dual of $\boldsymbol{D}$). In this way, we define the Fourier transform of any distribution as an element of $\boldsymbol{D}'$. The search for an elementary solution is thus transformed into the problem of division by a polynomial. The heuristic argument is as follows: We know that we can divide by a polynomial where the polynomial is large. Also, a polynomial cannot have "too many zeros," hence cannot be small too often. Now, the space $\boldsymbol{D}$ consists of entire functions; we make our division where the poly-

---

* Received February 16, 1954; revised March 18, 1954.

[1] Work partially supported by National Science Foundation Contract NSF 5-G205.

nomial is large and then use the maximum modulus theorem to show that things don't get too bad where the polynomial is small.

We are now preparing several papers in connection with the material presented here. In the first we shall study in detail the Fourier transform on the space $\boldsymbol{D}'$, and the structure of the elements of $\boldsymbol{D}'$. In another paper we shall study the structure of the elementary solutions in regard to regularity and growth at infinity. For example, we shall show that, for P any polynomial of derivation with constant coefficients, and a any positive number, we can find an elementary solution of P which, in some sense, increases less rapidly than $e^{a|x|}$. Moreover, if the highest derivative that occurs in P is of order k, and if P is a partial differential operator of n variables, then there exists an elementary solution which is no worse than a derivative of order $n+k+3$ of a continuous function. In another paper, we shall solve other problems in division. For example, we shall show that, if S is a distribution whose Fourier transform is an entire function of finite order, then division by S in the sense of convolution is possible in a certain space of distributions.

The methods of our paper, as well as most of the results of Sections I-V were obtained in April 1953 and were announced at that time in reports to the National Science Foundation. Since then, Malgrange [3], [3a] has obtained the corollary to Theorem 5 by a method which also uses the theory of analytic functions. Malgrange has shown how Theorem 5 implies the existence of an elementary solution and even the division property in $\mathcal{D}_F'$ (space of distributions of finite order). We include in Section IV a brief account of this part of Malgrange's work (Theorem 6).

Malgrange has also obtained our Theorem 10 by a different method. However, he does not obtain Theorem 9 which is apparently more powerful.

The author wishes to express his thanks to the referee for acquainting him with the work of Malgrange, as well as for his other useful comments.

2. Preliminaries. Let n be a positive integer which will be fixed throughout this paper. By C we denote the complex Euclidean n-space, and by R we denote the real part of C. For $z \in \binom{C}{R}$, we shall often write $z = (z_1, z_2, \cdots, z_n)$ where the z_j are $\binom{\text{complex}}{\text{real}}$ numbers. For $i = 1, 2, \cdots, 2^n$, let p_i be a point in C, $p_i = (p_{i_1}, p_{i_2}, \cdots, p_{i_n})$ where each $p_{i_j} = \pm 1$. (For convenience we denote by $p_1 = (1, 1, \cdots, 1)$.) Then by C^i we denote the subspace of C consisting of those points $z = (z_1, z_2, \cdots, z_n)$ such that, for all j, $\mathfrak{I}(z_j) = 0$, or $\operatorname{sgn} \mathfrak{I}(z_j) = \operatorname{sgn} p_j$ where, for a a complex number, $\mathfrak{R}(a)$ denotes the real part of a, and $\mathfrak{I}(a)$ denotes the imaginary part of a, and for b a real number $\neq 0$, $\operatorname{sgn} b$ is the signum of b.

For k a positive number, we shall denote by C_k the strip of width k around R in C, i. e. the set of all $z = (z_1, z_2, \cdots, z_n)$ with $|\Im(z_j)| \leqq k$ for all j.

Let F be a function on $\begin{pmatrix} C \\ R \end{pmatrix}$; let $r \leqq n$; let $j_1, j_2, \cdots, j_r$ be a subset of $(1, 2, \cdots, n)$, and let $a_{j_1} = a_{j_2}, \cdots, a_{j_r}$ be $\begin{pmatrix} \text{complex} \\ \text{real} \end{pmatrix}$ numbers. Then by $F_{X_{j_1} = a_{j_1}, X_{j_2} = a_{j_2}, \cdots, X_{j_r} = a_{j_r}}$ we denote the function on $\begin{pmatrix} \text{complex} \\ \text{real} \end{pmatrix}$ Euclidean $n - r$ space whose value at any point $(x_{i_1}, x_{i_2}, x_{i_{n-r}})$, where $((i_1, i_2, \cdots, i_{n-r})$ are those indices which do not appear in $(j_1, \cdots, j_r))$, is $F(w_1, w_2, \cdots, w_n)$ where $w_k = a_{j_p}$ if $k = j_p$, or $w_k = x_{i_p}$ if $k = i_p$.

We shall review, rapidly, those properties of the theory of distributions of Schwartz [5] which we shall need in our paper. For convenience, we shall depart slightly from the notation of Schwartz.

For l a positive number, denote by R_l the closed cube in R, center at the origin, and side $2l$. By $\mathfrak{D}_l$ we denote the space of indefinitely differentiable functions on R whose carriers are contained in R_l. (The carrier of a function is the closure of the set of points where it is different from zero.) A sequence of functions f_i converges to zero in $\mathfrak{D}_l$ if for P any partial differential operator on R with constant coefficients, $Pf_i \to 0$ uniformly on R. It is known that $\mathfrak{D}_l$ is a complete, metrizable topological vector space. We shall, unless explicitly stated to the contrary, use "vector space" to mean "vector space over the complex numbers."

By choosing a sequence of numbers l_i with $l_i \to \infty$, the spaces $\mathfrak{D}_{l_i}$ form a sequence of definition for a space $\mathfrak{D}$ of type $(\mathcal{LF})$ in the sense of Dieudonné and Schwartz [2]. $\mathfrak{D}$ is a complete topological Hausdorff space which is not metrizable. A function f on R is in $\mathfrak{D}$ if and only if f lies in some $\mathfrak{D}_l$, i. e. if and only if f is an indefinitely differentiable function with compact carrier. A fundamental system of neighborhoods of zero in $\mathfrak{D}$ consists of those convex sets N for which $N \cap \mathfrak{D}_l$ is a neighborhood of zero in $\mathfrak{D}_l$ for all l.

By $\mathfrak{D}'$ we denote the dual of $\mathfrak{D}$, that is, the space of continuous linear functions on $\mathfrak{D}$ with the topology of uniform convergence on the bounded sets of $\mathfrak{D}$. Unless explicitly stated otherwise, "linear" will be taken to mean "linear over the complex numbers." The important point to note is: Let S be a linear function on D whose restriction to $\mathfrak{D}_l$ is continuous on $\mathfrak{D}_l$ for each l; then $S \in \mathfrak{D}'$. Thus, to verify that a linear function on $\mathfrak{D}$ is continuous, we need show only that, for any l, if $\{f_i\}$ is a sequence of functions in $\mathfrak{D}_l$ which converge to zero in $\mathfrak{D}_l$, then $S(f_i) \to 0$.

If P is a partial differential operator with constant coefficients, (we sometimes say "P is a polynomial of derivation" if this is the case) then P clearly induces a continuous linear map of $\mathfrak{D}$ into $\mathfrak{D}$ defined by $f \to Pf$ for any $f \in \mathfrak{D}$. We shall denote by P' the adjoint of P, that is, for any $S \in \mathfrak{D}'$, $f \in \mathfrak{D}$, $P'S(f) = S(Pf)$. It is easily verified that P' is a continuous linear map of $\mathfrak{D}' \to \mathfrak{D}'$.

By "dx" we shall denote the ordinary Lebesgue measure on R divided by $(2\pi)^{n/2}$. Then for any $f \in \mathfrak{D}$, we define the Fourier transform F of f as the function on C, $F(z) = \int f(x) e^{-ix \cdot z} dx$, where for $z \in C$ and $x \in R$, $x \cdot z = x_1 z_1 + x_2 z_2 + \cdots + x_n z_n$. (When we omit the region of integration, it is to be assumed that this region is R.) F is an entire function of exponential type, and if $f \in \mathfrak{D}_l$, F is of exponential type $\leqq l$. Moreover, F is rapidly decreasing on R, i. e. for Q any polynomial on C QF is bounded on R. On the other hand, if F is an entire function of exponential type $\leqq l$ which is rapidly decreasing, then the function f on R defined by

$$f(x) = \int_R F(y) e^{ix \cdot y} dy,$$

is in $\mathfrak{D}$. Moreover, Fourier's inversion formula holds: for $f \in \mathfrak{D}$,

$$\text{if } F(z) = \int f(x) e^{-ix \cdot z} dx, \text{ then } f(x) = \int_R F(y) e^{ix \cdot y} dy.$$

We shall make the following convention of notation: Lower case letters will be used to denote functions of $\mathfrak{D}$ and the corresponding upper case letters will be used to denote their Fourier transforms.

Let P be a partial differential operator with constant coefficients on $\mathfrak{D}$. Then for $f \in \mathfrak{D}$, there is a polynomial Q on C such that

$$\int (Pf)(x) e^{-ix \cdot z} dx = Q(z) F(z).$$

In this case, we write $Q = \boldsymbol{F}(P)$. Conversely, given any polynomial Q on C, we can find a polynomial of derivation P with $Q = \boldsymbol{F}(P)$.

3. The space $\boldsymbol{D}$. For any $l > 0$, we denote by $\boldsymbol{D}_l$ the set of Fourier transforms of $\mathfrak{D}_l$. By what we said above, $\boldsymbol{D}_l$ consists of all entire functions of exponential type $\leqq l$ which are capidly decreasing on R. We shall define in $\boldsymbol{D}_l$ a topology to make the Fourier transform $\mathfrak{F}$ a topological isomorphism of $\mathfrak{D}_l$ onto $\boldsymbol{D}_l$. The following extension to n dimensions of a result of Phragmén and Lindelöf (see [6]) is needed.

LEMMA 1. *Let F be an entire function of exponential type $\leqq l$ which satisfies $|F(x)| \leqq M$ for all $x \in R$. Then for all $z \in C^j$,*

$$|F(z)\exp(ilp_j \cdot z)| \leqq M.$$

Proof. We shall adjust our notation to the case $j = 1$, the other cases being handled similarly. We use induction on n. For $n = 1$ the result is the known theorem of Phragmén and Lindelöf. Assume $n > 1$ and that the result is valid for $n - 1$. Let $a_1 \in R$; consider $F_{X_1=a_1}$. It is clear from the definitions that

$$\sup_{z \in R_{(n-1)}} |F_{X_1=a_1}(z)| \leqq M,$$

where $R_{(n-1)}$ denotes the real Euclidean $n-1$ space. Thus, for $C^1{}_{(n-1)}$, the analogue of C^1 for complex Euclidean $n-1$ space, by our induction assumption,

$$\sup_{z \in C^1{}_{(n-1)}} |[\exp il(z_2 + \cdots + z_n)]F_{X_1=a_1}(z)| \leqq M.$$

Now, let $(z_2, \cdots, z_n)$ be any point of $C^1{}_{(n-1)}$, and consider the function of one complex variable $z \to [\exp il(z_2 + \cdots + z_n)]F_{X_2=z_2, \cdots, X_n=z_n}(z)$. By the above,

$$\sup_{z \in R_{(1)}} |[\exp il(z_2 + \cdots + z_n)]F_{X_2=z_2, \cdots, X_n=z_n}(z)| \leqq M.$$

Thus, by the result for $n = 1$,

$$\sup_{z_1 \in C^1{}_{(1)}} |[\exp il(z_1 + z_2 + \cdots + z_n)]F_{X_2=z_2, \cdots, X_n=z_n}(z)| \leqq M,$$

or, what is the same thing,

$$\sup_{z_1 \in C^1{}_{(1)}} |[\exp il(z_1 + z_2 + \cdots + z_n)]F(z_1, z_2, \cdots, z_n)| \leqq M.$$

Since this holds for all $(z_2, \cdots, z_n) \in C^1{}_{(n-1)}$, we have

$$\sup_{z \in C^1} |[\exp il(z_1 + z_2 + \cdots + z_n)]F(z)| \leqq M,$$

which is the desired result.

From this lemma it follows that, for any $a > 0$, any polynomial P on C, and any function $f \in \boldsymbol{D}$, PF is bounded in C_a.

THEOREM 1. *Let $\{f_i\}$ be a sequence of functions of $\mathfrak{D}_l$. A necessary and sufficient condition that $f_i \to 0$ in $\mathfrak{D}_l$ is, for any $k > 0$ and any polynomial Q on C, $QF_i \to 0$ uniformly on C_k. It is sufficient to have, for some $k > 0$ and every polynomial P on C, $PF_i \to 0$ uniformly on C_k.*

Proof. a. *Sufficiency.* Assume that, for some $k > 0$ and any polynomial

Q on C, $QF_i \to 0$ uniformly on C_k. It follows immediately from Cauchy's formula for the derivatives of an analytic function of several complex variables, that, for D any operator of partial differentiation with constant coefficients, and Q any polynomial on C, $D(QF_i) \to 0$ uniformly on R. Thus, $F_i \to 0$ in the space $\mathcal{S}$ of Schwartz. It follows that $f_i \to 0$ in $\mathcal{S}$ (see [5]). Since the f_i all have their carriers in R_l, this means that $f_i \to 0$ in $\mathfrak{D}_l$.

b. *Necessity.* Assume $f_i \to 0$ in $\mathfrak{D}_l$; let Q be a polynomial on R, and $k > 0$. We know (see [5]) that $F_i \to 0$ in $\mathcal{S}$, and that the F_i are all of exponential type $\leqq l$. Thus, by the definition of the topology of $\mathcal{S}$, $QF_i \to 0$ uniformly on R; by Lemma 1, for any S with $1 \leqq S \leqq 2^n$,

$$| Q(z)F_j(z) \exp ilp_S \cdot z | \to 0,$$

uniformly for $z \in C^S$, *a fortiori* for $z \in C^S \cap C_k$. Now,

$$\min_{z \in C^S \cap C_k} | \exp ilp_S \cdot z | = e^{-lkn}.$$

Thus, $QF_j \to 0$ uniformly on $C^S \cap C_k$ for all S; this means that $QF_j \to 0$ uniformly on C_k.

For any positive number k and any polynomial Q in n variables, we define the function $v_{k,Q}$ on $\boldsymbol{D}_l$ by

$$v_{k,Q}(F) = \sup_{z \in C_k} | Q(z)F(z) |,$$

for any $F \in \boldsymbol{D}_l$; it is clear that the $v_{k,Q}$ are semi-norms. We use these $v_{k,Q}$ to define a topology on $\boldsymbol{D}_l$; we denote the space obtained again by $\boldsymbol{D}_l$, so that $\boldsymbol{D}_l$ is a locally convex topological vector space. If $F \in \boldsymbol{D}_l$, $F \neq 0$, then there is a $z_0 \in R$ with $F(z_0) \neq 0$. Thus,

$$v_{1,1}(F) = \sup_{z \in C_1} | F(z) | \geqq F(z_0) > 0,$$

so that $\boldsymbol{D}_l$ is Hausdorff. If Q is any polynomial, and k a positive number, there are integers $A > 0$, $m > 0$ so that, for all $z \in C_k$,

$$| Q(z) | \leqq | A + [\sum_{i=1}^{n} z_i^2]^m |.$$

The totality of polynomials of the form $A + [\sum_{i=1}^{n} z_i^2]^m$ is certainly a denumerable set T. It is clear that the topology of $\boldsymbol{D}_l$ can be defined by means of semi-norms $v_{k,S}$ for k rational and $S \in T$. We need thus only a denumerable number of semi-norms to define the topology of $\boldsymbol{D}_l$, which means that this space is metrizable.

Theorem 1 can now be restated as

THEOREM 2. *The Fourier transform* $\mathfrak{F}$ *is a topological isomorphism of* $\mathfrak{D}_l$ *onto* $\boldsymbol{D}_l$ *for each* $l > 0$.

Since each $\mathfrak{D}_l$ is complete, the same is true of $\boldsymbol{D}_l$, and these spaces are of type $(\mathfrak{F})$ (see [2]). It is clear that, if $i > j$, the topology induced by $\boldsymbol{D}_i$ on the set of functions of $\boldsymbol{D}_j$ is just the topology of the space $\boldsymbol{D}_j$. We may thus use the spaces $\boldsymbol{D}_i$ for i integral as a sequence of definition for a space $\boldsymbol{D}$, the inductive limit of the spaces $\boldsymbol{D}_i$ in the sense of Dieudonné and Schwartz [2]. Set theoretically, $\boldsymbol{D} = \bigcup_{i=1}^{\infty} \boldsymbol{D}_i$, i. e. a function F on C is in $\boldsymbol{D}$ if and only if F is an entire function of exponential type which has the property that QF is a bounded function on R for every polynomial Q in n indeterminates. The topology of $\boldsymbol{D}$ is defined as follows: A fundamental system of neighborhoods of zero in $\boldsymbol{D}$ consists of all convex sets N such that, for each $l > 0$, $N \cap \boldsymbol{D}_l$ is a neighborhood of zero in $\boldsymbol{D}_l$. $\boldsymbol{D}$ is locally convex and complete, but not metrizable. From our Theorem 2 and Proposition 5 of [2] we obtain

THEOREM 3. *The Fourier transform* $\mathfrak{F}$ *is a topological isomorphism of* $\mathfrak{D}$ *onto* $\boldsymbol{D}$.

By $\boldsymbol{D}'$ we denote the dual of $\boldsymbol{D}$ with the topology of uniform convergence on the bounded sets of $\boldsymbol{D}$. The important thing to know about $\boldsymbol{D}'$ is: (see [2] Proposition 5) Let S be a linear function on $\boldsymbol{D}$ which is continuous on each $\boldsymbol{D}_l$; then S is continuous on $\boldsymbol{D}$, i. e. $S \in \boldsymbol{D}'$. Since each $\boldsymbol{D}_l$ is metrizable, in order to verify that S is continuous on $\boldsymbol{D}_l$, we need consider only *sequences.* For sequences, Lemma 1 gives us an explicit description of convergence. It is by means of this explicit description that we are able to show the existence of an elementary solution.

Let us denote by $\boldsymbol{F}$ the adjoint of $\mathfrak{F}^{-1}$, so that $\boldsymbol{F}$ is the map of $\mathfrak{D}' \to \boldsymbol{D}'$ defined by $(\boldsymbol{F}S)(F) = S(f)$, for any $S \in \mathfrak{D}'$, $f \in \mathfrak{D}$. It is natural to call $\boldsymbol{F}$ the Fourier transform on $\mathfrak{D}'$. (This will be discussed in a forthcoming article.) Thus, the Fourier transform of any distribution, in particular, of any locally summable function, is defined. From Theorem 3 we obtain easily

THEOREM 4. $\boldsymbol{F}$ *is a topological isomorphism of* $\mathfrak{D}'$ *onto* $\boldsymbol{D}'$.

We want to compute $\boldsymbol{F}(\delta)$ where δ is the Dirac measure, that is, the measure consisting of a mass of $+1$ placed at the origin. We have

$$(\boldsymbol{F}\delta)(F) = \delta(f) = f(0) = \int F(x)\,dx,$$

so that $\mathfrak{F}\delta = 1$ where 1 is the element of $\boldsymbol{D}'$, $1(F) = \int F(x)\,dx$, for all $F \in \boldsymbol{D}$.

4. Existence of an elementary solution.

LEMMA 2. *Let there be given m points $x_1, x_2, \cdots, x_m$ in the complex plane. Then about any complex number a we can draw a circle L of radius $\leqq 1$ so that, for any $x \in L$, $\min |x - x_i| \geqq 1/2(m+1)$.*

Proof. We may certainly restrict ourselves to the case where a is the origin. Let $x_1, \cdots, x_r$ lie in or on the unit circle and $x_{r+1}, \cdots, x_m$ lie outside the unit circle. Then we arrange the $\leqq r+2$ points $0, |x_1|, \cdots, |x_r|, 1$ on the unit interval I. We have thus decomposed I into at most $r+1$ subintervals; it follows that one of these subintervals is of length at least $1/r+1$. Let b be the midpoint of this interval, and let L be the circle, center origin, and radius b. Then for any $x \in L$, $i = 1, 2, \cdots, n$.

$$|x - x_i| \geqq |x| - |x_i| = |(b - |x_i|)| \geqq 1/2(r+1) \geqq 1/2(m+1).$$

LEMMA 3. *Let Q be a polynomial in one indeterminate,*

$$Q(x) = Q_0 X^m + Q_1 X^{m-1} + \cdots + Q_m.$$

Then, there is a constant $\alpha \neq 0$ (depending only on m) so that, about any complex number a we can draw a circle M of radius $\leqq 1$ for which

$$|Q(x)| \geqq \alpha |Q_0|$$

for all $x \in M$. (It is allowed that $Q_0 = 0$.)

Proof. For $Q_0 = 0$, the result is trivial. Let $Q_0 \neq 0$ and write

$$Q(X) = Q_0(X - x_1)(X - x_2) \cdots (X - x_m),$$

where $x_1, \cdots, x_m$ are the zeros of Q. Then, by Lemma 2, we can draw about a a circle M of radius $\leqq 1$ so that $|x - x_i| \geqq 1/2(m+1)$, for all $x \in M$, $1 \leqq i \leqq m$. Thus, for $x \in M$, $|Q(x)| \geqq |Q_0| 1/2^m (m+1)^m$. The result follows by choosing $\alpha = 1/2^m (m+1)^m$.

The main lemma used for the proof of the existence of an elementary solution is the following:

LEMMA 4. *Let Q be a non-constant polynomial in the letters $X_1, \cdots, X_n$. Let $r > 1$ and $F \in \boldsymbol{D}$ satisfy*

$$\sup_{z \in C_r} |Q(z)F(z)| \leqq a.$$

Suppose X_p occurs in Q, and write $Q = Q_0 X_p^m + \cdots + Q_m$, where the Q_i are polynomials in $X_1, \cdots, X_{p-1}, X_{p+1}, \cdots, X_n$. Then we can find a $\beta > 0$ such that

$$\sup_{z \in C_{r-1}} |Q_0(z)F(z)| \leqq a\beta.$$

Proof. We may suppose without loss of generality that $p = 1$. Let $z_1, z_2, \cdots, z_n$ be complex numbers with $|\mathfrak{J}(z_1)| \leqq r-1, \cdots, |\mathfrak{J}(z_n)| \leqq r-1$. By Lemma 3, we can find a number $\alpha > 0$, and a circle M with centers z_1 and radius $\leqq 1$ so that, for all $z \in M$,

$$|Q_{X_2=z_2, \cdots, X_n=z_n}(z)| \geqq \alpha |Q_0(z_2, \cdots, z_n)|.$$

Thus, for $z \in M$,

$$\begin{aligned}\alpha |Q_0(z_2, \cdots, z_n)F_{X_2=z_2, \cdots, X_n=z_n}(z)| &\leqq |Q_{X_2=z_2, \cdots, X_n=z_n}(z)F_{X_2=z_2, \cdots, X_n=z_n}(z)| \\ &= |Q(z, z_2, \cdots, z_n)F(z, z_2, \cdots, z_n)| \leqq a,\end{aligned}$$

because $(z, z_2, \cdots, z_n) \in C_r$. On choosing $\beta = 1/\alpha$, we find

$$|Q_0(z_2, \cdots, z_n)F_{X_2=z_2, \cdots, X_n=z_n}(z)| \leqq a\beta,$$

for all $z \in M$.

We have two cases: If $Q_0(z_2, \cdots, z_n) = 0$, we certainly have

$$|Q_0(z_2, \cdots, z_n)F_{X_2=z_2, \cdots, X_n=z_n}(z_1)| \leqq a\beta.$$

If $Q_0(z_2, \cdots, z_n) \neq 0$, then for all $z \in M$

$$|F_{X_2=z_2, \cdots, X_n=z_n}(z)| \leqq a\beta / |Q_0(z_2, \cdots, z_n)|.$$

Since $F_{X_2=z_2, \cdots, X_n=z_n}$ is analytic, we get, by the maximum modulus theorem,

$$|F_{X_2=z_2, \cdots, X_n=z_n}(z_1)| \leqq a\beta / Q_0(z_2, \cdots, z_n).$$

Thus, in any case, $|Q_0(z_2, \cdots, z_n)F(z_1, \cdots, z_n)| \leqq a\beta$. Since $(z_1, \cdots, z_n)$ was an arbitrary point of C_{r-1}, the lemma is established. We can now state our main result:

THEOREM 5. *Let Q be a polynomial different from zero in n indeterminates, and let $\{F_i\}$ be a sequence of functions in $\boldsymbol{D}$. Suppose that $QF_i \to 0$ in $\boldsymbol{D}$; then $F_i \to 0$ in $\boldsymbol{D}$.*

Proof. We use induction on the number k of variables which occurs in Q. For $k = 0$, Q is a constant $\neq 0$ and the result is trivial. Assume then that $k > 0$ and that the result is proven for all polynomials in $k-1$ indeterminates. We may assume without loss of generality that $X_1, \cdots, X_k$ actually occur in Q and write

$$Q(X_1, \cdots, X_k) = Q_0(X_2, \cdots, X_k)X_1^m + \cdots + Q_m(X_0, \cdots, X_k),$$

where $Q_0 \neq 0$.

Under our hypotheses, we know (see [2]) that the functions QF_i all lie in some $\mathcal{D}_l$ and that $QF_i \to 0$ in $\mathcal{D}_l$. Since the exponential type of QF_i is the same as that of F_i, the F_i also lie in $\boldsymbol{D}_l$. Theorem 1 tells us that, for any $r > 0$ and any polynomial P on C, $PQF_i \to 0$ uniformly on C_r. Lemma 4 shows us that the same is true for the sequence $\{Q_0F_i\}$, that is, for any $r > 0$ and any polynomial P in n indeterminates, $Q_0PF_i \to 0$ uniformly on C_r. This means that $Q_0F_i \to 0$ in $\boldsymbol{D}_l$. By our induction hypothesis, $F_i \to 0$ in $\boldsymbol{D}_l$ which implies $F_i \to 0$ in $\boldsymbol{D}$.

COROLLARY. *Let P be an operator of partial differentiation with constant coefficients on $\mathcal{D}$. Suppose $\{f_i\}$ is a sequence of functions of $\mathcal{D}$ such that $Pf_i \to 0$ in $\mathcal{D}$. Then also $f_i \to 0$ in $\mathcal{D}$.*

For the sake of completeness, we shall show, briefly, how Malgrange uses this corollary (or rather, a similar result) to obtain elementary solutions, and division in a certain space (see [3b]).

For any $l > 0$, $s \geqq 0$, denote by $\mathcal{D}_l^s$ the space of functions on R which are s times continuously differentiable and which have their carriers in K_l. The topology of $\mathcal{D}_l^s$ is defined by means of the semi-norms $v_D(f) = \max_{x \in R} |(Df)(x)|$ for $f \in D_l^s$ and D any partial differential operator with constant coefficients of order $\leqq s$. By $\mathcal{D}^s$ we denote the inductive limit of the $\mathcal{D}_l^s$ for l integral. $\mathcal{D}'^s$ denotes the dual of $\mathcal{D}^s$; this is the space of distributions of order s. We write $\mathcal{D}_F' = \bigcup_{s=1,2,\cdots} \mathcal{D}'^s$ so $\mathcal{D}_F'$ is the space of distributions of finite order.

Suppose P is a partial differential operator with constant coefficients of order k. Then by $P\mathcal{D}_l^{s+k}$ we denote the subset of $\mathcal{D}_l^s$ consisting of all functions of the form Pf for $f \in \mathcal{D}_l^{s+k}$; the topology of $P\mathcal{D}_l^{s+k}$ is that induced by $\mathcal{D}_l^s$.

THEOREM 6. *For any $S \in \mathcal{D}_F'$, there is a $T \in \mathcal{D}_F'$ with $P'T = S$. In particular, P has an elementary solution in $\mathcal{D}_F'$.*

Proof. Suppose that S is a distribution of order s. By reasoning analogous to that in the proofs of Theorem 5 and its corollary, we can show that $Pf \to f$ is a continuous linear map of $P\mathcal{D}^{s+k+3} \to \mathcal{D}^s$. Thus, $\tilde{T}: Pf \to S \cdot f$ is a continuous linear function on $P\mathcal{D}^{s+k+3}$. Let K be the set of functions $g \in P\mathcal{D}^{s+4+k}$ for which

$$\max_{x \in R} \left| \frac{\partial^{p_1+p_2+\cdots+p_n}}{\partial X_1^{p_1} \partial X_2^{p_2} \cdots \partial X_n^{p_n}} g(x) \right| \leqq 1,$$

for all $p_1, \cdots, p_n$ such that $p_1 + \cdots + p_n \leqq s + 4$. Then, for each integer l, $K_l = K \cap \mathcal{D}_l^{s+3}$ is a compact set in $P\mathcal{D}_l^{s+3+k}$ by the theorem of Ascoli (see [5]). For each integer l, let U_l be the set of $f \in K_l$ which satisfy $|T \cdot f| < 1$,

and call v the convex closure of $\cup U_l$. Then, by Proposition 1 of [3b], v is a neighborhood of zero in $P\mathcal{D}^{s+k+4}$. Further, for $f \in v$, $|\tilde{T} \cdot f| \leqq 1$. Thus, $\tilde{T}$ is continuous on $P\mathcal{D}^{s+k+4}$ and hence can be extended to a continuous linear function T on $\mathcal{D}^{s+4}$ by the Hahn-Banach theorem. For any $f \in \mathcal{D}$, $P'T \cdot f = T \cdot Pf = T \cdot Pf = S \cdot f$. Thus, $P'T = S$; if we take $S = \delta$, T is the elementary solution for P.

5. Division in the space $\mathcal{D}_F'$. We shall give an independent proof of Theorem 6 by means of the Fourier transform, without using the notions of Malgrange.

We introduce a topology in $\mathcal{D}_F = \cap \mathcal{D}^s$ by saying that a set N in $\mathcal{D}_F$ is a neighborhood of zero if it is the intersection with $\mathcal{D}_F$ of a neighborhood of zero in some $\mathcal{D}^s$. (The elements of $\mathcal{D}_F$ are the same as those of $\mathcal{D}$, but the topology of $\mathcal{D}$ is stronger than that of $\mathcal{D}_F$.) It follows immediately from the Hahn-Banach theorem that $\mathcal{D}_F'$ is the dual of $\mathcal{D}_F$. By $\boldsymbol{D}_F$ we denote the space of Fourier transforms of functions of $\mathcal{D}_F$ with the topology to make $\mathcal{F} : \mathcal{D}_F \to \boldsymbol{D}_F$ a topological isomorphism. $\boldsymbol{D}_F$ consists, of course, of all rapidly decreasing entire functions of exponential type.

THEOREM 7. *The topology of $\boldsymbol{D}_F$ can be described as follows: A fundamental system of neighborhoods of zero in $\boldsymbol{D}_F$ consists of all sets $N \subset \boldsymbol{D}_F$ for which we can find a finite sequence of polynomials $A_1, A_2, \cdots, A_r$ and an increasing sequence of positive numbers $a_l \to \infty$ so that N consists of all $G \in \boldsymbol{D}_F$ which satisfy, for each integer l,*

$$\max_{z \in C^j \cap C_{a_l}} |(\exp ilp_j \cdot z) A_k(z) G(z)| \leqq 1, \tag{5.1}$$

for $k = 1, 2, \cdots, r$, and $j = 1, 2, \cdots, n$.

Proof. Suppose first that N satisfies the conditions stated in the theorem. Then N is clearly convex, and for any integer $m > 0$ and any $G \in \boldsymbol{D}_m$, the conditions

$$\max_{z \in C^j \cap C_{a_l}} |(\exp ilp_j \cdot z) A_k(z) G(z)| \leqq 1,$$

for any k, j, and $l > m$ are already implied by the conditions

$$\max_{z \in C^j \cap C_{a_m}} |(\exp imp_j \cdot z) A_k(z) G(z)| \leqq 1,$$

for $k = 1, 2, \cdots, r$, and $j = 1, 2, \cdots, n$ because of the Phragmén-Lindelöf theorem (Lemma 1). Thus, by the definition of the topology of $\boldsymbol{D}_m$, $N \cap \boldsymbol{D}_m$ is a neighborhood of zero in $\boldsymbol{D}_m$. It is clear from the proof of Theorem 1,

that the fact that we are using a fixed set of polynomials $A_1, \cdots, A_r$ for the definition of N means that we can find an s so that, for all m, $N' \cap \mathfrak{D}_m$ is a neighborhood of zero in $\mathfrak{D}_m{}^s$, where N' is the set of inverse Fourier transforms of the functions of N. Thus, by the definition of the topology of $\boldsymbol{D}_F$, N is a neighborhood of zero in $\boldsymbol{D}_F$.

Conversely, let M be a neighborhood of zero in $\boldsymbol{D}_F$, and denote by M' the set of inverse Fourier transforms of functions of M. Then, we can find a finite set of polynomials of derivation $B_1, B_2, \cdots, B_t$ and a decreasing sequence of numbers $b_l \to 0$ so that M' contains the set of all $f \in \mathfrak{D}$ which satisfy

$$\max_{x \notin K_{(n+1)l}} |(B_k f)(x)| \leqq b_l, \tag{5.2}$$

for all l and $k = 1, 2, \cdots, t$, and also

$$\max_{x \in K_{n+1}} |(B_k f)(x)| \leqq 1, \tag{(5.3)}$$

for $k = 1, 2, \cdots, t$. (This can be shown exactly as the analogous description of the topology of $\mathfrak{D}$ (see [5]) is demonstrated.)

Let us denote by $A_1, A_2, \cdots, A_r$ the polynomials $4n$, $4n\boldsymbol{F}(B_k)$, $4nz_j{}^2\boldsymbol{F}(B_k)$, for $k = 1, 2, \cdots, t$ and $j = 1, 2, \cdots, n$. We define the sequence $\{a_l\}$ by $a_l = \max(1, -\log b_l)$. Let N be the set of $G \in \boldsymbol{D}_F$ which satisfy, for each l,

$$\max_{z \in C^j \cap C_{a_l}} |(\exp ilp_j \cdot z) A_k(z) G(z)| \leqq 1, \tag{5.4}$$

for $k = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, n$; let N' denote the set of inverse Fourier transforms of N. We claim that $N' \subset M'$.

It is clear that any $g \in N'$ satisfies (5.3). Let $g \in N'$ and $x \notin K_{(n+1)l}$. Let j be chosen so that (see Section 2) $\operatorname{sgn} p_{j_k} = \operatorname{sgn} x_k$ for $k = 1, 2, \cdots, n$. It is clear from Caucy's theorem that we may write $g(x) = \int_{\alpha_l} G(z) \exp(ix \cdot z) dz$, where α_l is the plane $\mathfrak{I}(z_k) = a_l p_{j_k}$ for $k = 1, 2, \cdots, n$. Thus,

$$g(x) = \int_{\alpha_l} G(z) \exp(ilp_j \cdot z) \exp(i(x - lp_j) \cdot z) dz.$$

Now, since $x \notin K_{(n+1)l}$, for at least one k, $x_k > (n+1)l$. Thus, for $z \in \alpha_l$, $\mathfrak{I}[(x - lp_j) \cdot z] \geqq (nl)a_l - (n-1)la_l = la_l$. This means that, for $z \in \alpha_l$, $|\exp i(x - lp_j) \cdot z| \leqq e^{-la_l}$. Moreover, it is clear from (5.4) that

$$\int_{\alpha_l} |G(z)(\exp ilp_j \cdot z)| \, dz \leqq 1.$$

Thus, $|g(x)| \leqq e^{-a_l} \leqq b_l$.

A similar argument shows that, for $k = 1, 2, \cdots, t$, $|(B_k g)(x)| \leqq b_i$. This means that $g \in M'$ which is the desired result.

Let Q be a polynomial $\neq 0$ in n letters. We denote by $Q\boldsymbol{D}_F$ the space of all QG for $G \in \boldsymbol{D}_F$ with the topology induced by $\boldsymbol{D}_F$.

THEOREM 8. *$QG \to G$ is a continuous linear map of $Q\boldsymbol{D}_F \to \boldsymbol{D}_F$.*

Proof. We use induction on the number k of letters that occur in Q. For $k = 0$ the result is obvious; assume the theorem is true for all polynomials with fewer than k letters, and let Q be a polynomial in k letters. We may assume that X_n occurs in Q and write $Q = Q_0 X^r + Q_1 X^{r-1} + \cdots + Q_r$, where the Q_j are polynomials in fewer than k letters. The result will follow if we can show that $QG \to Q_0 G$ is a continuous linear map of $Q\boldsymbol{D}_F \to Q_0\boldsymbol{D}_F$.

By linearity, we need prove continuity only at zero. Let then N be a neighborhood of zero in $Q_0\boldsymbol{D}_F$. By Theorem 7 we can find polynomials $A_1, A_2, \cdots, A_r$ and an increasing sequence of positive numbers a_l so that N contains all $G \in Q_0\boldsymbol{D}_F$ which satisfy, for each l,

$$\max_{z \in C^j \cap C_{a_l}} |(\exp ilp_j \cdot z) A_k(z) G(z)| \leqq 1$$

for $k = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, n$. Let β be a constant which satisfies the conclusions of Lemma 4. For each l, write $a_l' = a_l + 1$ and call $A_k' = 1/(e\beta) A_k$. Let N' be the set of $G \in QD_F$ which satisfy, for each l, k, j,

$$\max_{z \in C^j \cap C_{a_l}} |(\exp ilp_j \cdot z) A_k'(z) G(z)| \leqq 1.$$

Then, as in the proof of Lemma 4 we can show that, for $QG \in N'$, $Q_0 G \in N$ which concludes the proof of Theorem 8.

For P a polynomial of derivation, denote by $P\mathfrak{D}_F$ the space of Pf, $f \in \mathfrak{D}_F$ with the topology induced from $\mathfrak{D}_F$. Then we have an immediate

COROLLARY. *$Pf \to f$ is a continuous linear map of $P\mathfrak{D}_F \to \mathfrak{D}_F$.*

We can now furnish another proof of Theorem 6: Let $S \in \mathfrak{D}_F'$; then the map $\tilde{T}: Pf \to S \cdot f$ is a continuous linear function on $P\mathfrak{D}_F$ being the composition of the continuous maps $Pf \to f$ and $f \to S \cdot f$. Thus, by the Hahn-Banach theorem, T can be extended to a continuous linear function on $\mathfrak{D}_F$, i.e. to a $T \in \mathfrak{D}_F'$. We have, for $f \in \mathfrak{D}_F$, $P'T \cdot f = T \cdot Pf = \tilde{T} \cdot Pf = S \cdot f$. Thus, $P'T = S$ which is the desired result.

6. Division in the space $\mathcal{E}$. $\mathcal{E}$ is the space of indefinitely differentiable functions. $\mathcal{E}'$, its dual, is the space of distributions of compact carrier (see [5]). We shall use the following characterization of the space $\mathcal{E}'$ which will be established in our forthcoming paper, "Analytic Functions and the Fourier Transform of Distributions":

LEMMA 7. *$\mathcal{E}'$ consists exactly of those distributions $S \in \mathcal{D}'$ for which the convolution $S * f \in \mathcal{D}$ for any $f \in \mathcal{D}$. Moreover, the topology of $\mathcal{E}'$ is the same as that obtained by considering every $S \in \mathcal{E}'$ as defining the continuous linear transformation $f \to S * f$ of $\mathcal{D} \to \mathcal{D}$, and then giving this set of transformations the compact-open topology. That is, a fundamental system of neighborhoods of zero in $\mathcal{E}'$ consists of those sets N for which we can find a compact set $K \subset \mathcal{D}$ and a neighborhood N of zero in $\mathcal{D}$ so that N consists of all $S \in \mathcal{E}'$ for which $S * f \in M$ for all $f \in K$.*

Let us denote by $\mathcal{E}_F'$ the space of all $S \in \mathcal{D}_F'$ such that $S * f \in \mathcal{D}_F$ for all $f \in \mathcal{D}_F$; we consider each $S \in \mathcal{D}_F'$ as defining the continuous linear map $f \to S * f$ of $\mathcal{D}_F \to \mathcal{D}_F$, and we give the space $\mathcal{E}_F'$ the compact-open topology of this space of transformations. By Lemma 7, the elements of $\mathcal{E}_F'$ are the same as those of $\mathcal{E}'$. Moreover, if K is a compact set of $\mathcal{D}_F$, then K is clearly a compact set in each $\mathcal{D}^s$. Thus, by [2], K is contained in some $\mathcal{D}_L$. Since $\mathcal{D}_F$ and $\mathcal{D}$ induce the same topology on $\mathcal{D}_L$, K is compact in $\mathcal{D}$. Thus, the topology of $\mathcal{E}'$ is stronger than that of $\mathcal{E}_F'$.

We conclude from the above that, if we denote by $\mathcal{E}_F$ the dual of $\mathcal{E}_F'$, then for any $U \in \mathcal{E}_F$ we can find an $f \in \mathcal{E}$ such that, for all $S \in \mathcal{E}_F'$, $U \cdot S = S \cdot f$.

LEMMA 8. *For any $f \in \mathcal{E}$, $S \to S \cdot f$ is a continuous linear function on $\mathcal{E}_F'$.*

Proof. The linearity being obvious, we need prove continuity only at zero. For each lattice point $\gamma \in R$, let L_γ be the closed cube, center γ, side 4. Let $\{h_\gamma\}$ be a sequence of functions of $\mathcal{D}$ such that, for each γ, carrier $h_\gamma \subset L_\gamma$, and, for any $x \in R$, $\sum h_\gamma(x) = 1$ (this sum has meaning because only a finite number of terms are different from zero). For each γ, and $g \in \mathcal{E}$, denote by $\tau_\gamma g$ the function $(\tau_\gamma g)(x) = g(x - \gamma)$ for any $x \in R$. For each γ, choose $a_\gamma > 0$ such that $a_\gamma = a_{-\gamma}$ and

$$a_\gamma \geqq \max_{x \in R,\, 1 \leqq j \leqq i} |[Q_j(\tau_{\gamma_j}(h_{\gamma_j} f))](x)|,$$

where $\{Q_j\}$ is an enumeration of the polynomials of derivation on R with rational coefficients, and $\{\gamma_k\}$ is an enumeration of the lattice points of R, and where $\gamma = \gamma_i$.

Let K be the closure in $\mathcal{D}$ of the set of all functions $1/(a_\gamma)\tau_\gamma(h_\gamma f)$ so

that $K \subset \mathcal{D}_1$ and K is clearly bounded in $\mathcal{D}_1$. Thus (see [5]) K is compact in $\mathcal{D}$. Let N be the neighborhood of zero in $\mathcal{D}_F$ consisting of all $g \in \mathcal{D}_F$ such that, for each γ

$$\max_{x \in L_\gamma} |g(x)| \leqq b_\gamma / \max(1, a_\gamma).$$

Where $\{b_\gamma\}$ is a sequence of positive numbers with $\sum b_\gamma = 1$. Call M the set of $S \in \mathcal{E}_F'$ such that $S * p \in N$ for all $p \in K$; we claim that, for any $S \in M$, $|S \cdot f| \leqq 1$.

To this end, let $S \in M$, and suppose that the carrier of S is contained in K_l. Let r be chosen so that $\sum_{j=1}^{r} h_{\gamma_j}(x) = 1$ for any $x \in K_l$. Then $S \cdot f = \sum_{j=1}^{r} S \cdot h_j f$. But,

$$|S \cdot h_j f| = |(S * h_j f)(0)| = |(S * \tau_{\gamma_j}(h_j f))(\gamma_j)| \leqq b_{\gamma_j}.$$

Thus, $|S \cdot f| \leqq \sum_{j=1}^{r} b_m < 1$, which completes the proof of the lemma.

Let P be a constant-coefficient partial differential operator. By $P'\mathcal{E}_F'$ we denote the space of all $P'S$ for $S \in \mathcal{E}_F'$ with the topology induced by $\mathcal{E}_F'$. It follows immediately from the theorem of Paley and Wiener (see [5]) that P' is one-one. Our main result of this section is:

THEOREM 9. *$P'S \to S$ is a continuous linear map of $P'\mathcal{E}_F' \to \mathcal{E}_F'$.*

Proof. Since the map is clearly linear, it suffices to prove continuity at zero. Let $P'G$ be a filter base in $P'\mathcal{E}_F'$ which converges to zero in $P'\mathcal{E}_F'$ We denote by G the filter base obtained from $P'G$ by replacing each distribution $P'S$ of $P'G$ by S; the filter bases $P'G * f$ for $f \in \mathcal{D}$ are defined similarly. Then, if K is any compact set in $\mathcal{D}$, $P'G * f$ converge to zero in $\mathcal{D}_F$ uniformly for $f \in K$. Thus, $G * Pf$ converge to zero in $\mathcal{D}_F$ uniformly for $f \in K$. But, for any $S \in E'$, $S * Pf = P(S * f)$ (see [5]). Thus, $P(G * f)$ converge to zero in $\mathcal{D}_F$ uniformly for $f \in K$, which is the same as the uniform convergence of $P(G * f)$ in $P\mathcal{D}_F$. Now, $Pg \to g$ is a continuous linear map of $P\mathcal{D}_F \to \mathcal{D}_F$ by the corollary to Theorem 8. Hence $G * f$ converge to zero in $\mathcal{D}_F$ uniformly for $f \in K$. This means that G converges to zero in $\mathcal{E}_F'$ which is the desired result.[2]

We can now use this result to show that division is always possible in E:

[2] Footnote added in proof: In our forthcoming paper, "Analytic Functions and the Fourier Transform of Distributions," it will be shown that, in fact, the spaces $\mathcal{E}_F'$ and $\mathcal{E}'$ are the same.

THEOREM 10. *For any $f \in \mathcal{E}$ there is a $g \in \mathcal{E}$ with $Pg = f$.*

Proof. The map $\tilde{g}: P'S \to S \cdot f$ is a continuous linear function on $P'\mathcal{E}_F'$ because it is the composition of the continuous maps $P'S \to S$, $S \to S \cdot f$. Since $P'\mathcal{E}_F'$ is given the topology induced by $\mathcal{E}_F'$, $\tilde{g}$ can be extended to a continuous linear function $\hat{g}$ on $\mathcal{E}_F'$ by the Hahn-Banach theorem. By our above remarks, there is a $g \in \mathcal{E}$ such that $\hat{g} \cdot S = S \cdot g$ for all $S \in \mathcal{E}_F'$. Thus, for any $S \in \mathcal{E}_F$, $S \cdot Pg = P'S \cdot g = \hat{g} \cdot P'S = \tilde{g} \cdot P'S = S \cdot f$. Because of the Hahn-Banach Theorem, $Pg = f$.

7. Systems of equations. We shall extend the results of the previous section to systems of constant-coefficient partial differential equations. Let N be a positive number which will be fixed throughout this paper. By $\mathcal{D}^N$ we denote the space of N-tuples of functions of $\mathcal{D}$; the topology of $\mathcal{D}^N$ is that of coordinate-wise convergence. For any $l > 0$ we denote by $\mathcal{D}_l^N$ the space of N tuples of functions of $\mathcal{D}_l$ with topology of coordinatewise convergence. It is clear that $\mathcal{D}^N$ is the inductive limit of the spaces $\mathcal{D}_l^N$ for l integral. By $\mathcal{D}^{*N}$ we denote the set of continuous linear mappings of $\mathcal{D}^N$ into complex N-space. Thus, every element S of $\mathcal{D}^{*N}$ can be represented by an N by N matrix (S_{ij}) where each $S_{ij} \in \mathcal{D}'$ and where, for $f = (f_1, \cdots, f_N) \in \mathcal{D}^N$, $S(f) = (g_1, \cdots, g_N)$ with $g_i = \sum_{j=1}^{N} S_{ij} f$ for each i; the topology of $\mathcal{D}^{*N}$ is that of elementwise convergence of these matrices. (By "matrix" we shall mean "N by N matrix" unless explicitly stated otherwise.)

Let $\mathfrak{P} = (P_{ij})$ where the P_{ij} are polynomials of differentiation. Then $\mathfrak{P}$ defines a continuous map of $\mathcal{D}^N \to \mathcal{D}^N$ by $Pf = (g_1, \cdots, g_N)$ where, for each i, $g_i = \sum_{j=1}^{N} P_{ij} f_j$. By $\mathfrak{P}'$ we denote the adjoint of $\mathfrak{P}$, this is the map of $\mathcal{D}^{*N} \to \mathcal{D}^{*N}$ with $(\mathfrak{P}'S)(f) = S(\mathfrak{P}f)$, for all $S \in \mathcal{D}^{*N}$, $f \in \mathcal{D}^N$. By an elementary solution for $\mathfrak{P}$ we mean an $S \in \mathcal{D}^{*N}$ such that $\mathfrak{P}'S = \delta^N$, where

$$\delta^N = \begin{bmatrix} \delta & & 0 \\ & \ddots & \\ 0 & & \delta \end{bmatrix}$$

is the generalized Dirac measure.

For any $f \in \mathcal{D}^N$, $f = (f_1, \cdots, f_N)$, we define the Fourier transform of f: $F = \mathfrak{F}^N(f) = (F_1, \cdots, F_n)$. Thus, $\mathfrak{F}^N$ maps $\mathcal{D}^N$ onto $\boldsymbol{D}^N$ (space of N-tuples of functions of $\boldsymbol{D}$) and for any $l > 0$, $\mathfrak{F}^N$ maps $\mathcal{D}_l^N$ onto $\boldsymbol{D}_l^N$ (N-tuples of functions of $\boldsymbol{D}_l$). We give the spaces $\boldsymbol{D}^N$ and $\boldsymbol{D}_l^N$ the topology of coordinate-wise convergence and $\mathfrak{F}^N$ is clearly a topological isomorphism.

Thus, $\boldsymbol{D}^N$ is the inductive limit of the $\boldsymbol{D}_l$ for l integral. By $\boldsymbol{D}^{*N}$ we denote the set of linear transformations of $\boldsymbol{D}^N$ into complex N space. An element T of $\boldsymbol{D}^{*N}$ is thus a matrix (T_{ij}) with each $T_{ij} \in \boldsymbol{D}'$ and, for $F \in \boldsymbol{D}^N$,

$$T(F) = (\sum_{j=1}^{N} T_{ij}F, \cdots, \sum_{j=1}^{N} T_{Nj}F).$$

By $\boldsymbol{F}^N$ we denote the adjoint of $(\mathfrak{F}^N)^{-1}$, that is, $\boldsymbol{F}^N$ is the map of $\mathfrak{D}^{*N} \to \mathfrak{D}^{*N}$ defined by $(\boldsymbol{F}^N S)(G) = S(g)$, for any $S \in \mathfrak{D}^{*N}$, $g \in \mathfrak{D}^N$. (We again use the convention of representing an element of $\mathfrak{D}^N$ by a lower case letter, and using the same upper case letter for its Fourier transform.) For $f \in \mathfrak{D}^N$ we shall write $f = (f_1, \cdots, f_N)$ where all $f_i \in \mathfrak{D}$, with a similar convention for $\boldsymbol{D}^N$. $\boldsymbol{F}^N$ is again a topological isomorphism of $\mathfrak{D}^{*N} \to \boldsymbol{D}^{*N}$, which we can consider as defining the Fourier transform on $\mathfrak{D}^{*N}$.

Let $S = (S_{ij}) \in \mathfrak{D}^{*N}$. Then we can write

$$S = \sum_{k=1}^{N^2} T^{(k)}, \quad \text{where} \quad T_{ij}^{(k)} = \begin{cases} S_{ij} & \text{for } k = i + N(j-1). \\ 0 & \text{otherwise.} \end{cases}$$

We compute $\boldsymbol{F}^N(T^{(k)})$: For $F \in \boldsymbol{D}^N$, $\boldsymbol{F}^N(T^{(k)})(F) = T^{(k)}(f) = S_{ij}(f_j)$, where $k = i + N(j-1)$. Call $U_{ij} = \boldsymbol{F}(S_{ij})$; for $k = 1, 2, \cdots, N^2$, call $N^{(k)}$ the matrix for which $V_{ij}^{(k)}$ is U_{ij} or 0 according as k is or is not of the form $i + N(j-1)$. Then

$$\boldsymbol{F}^N(T^{(k)})(F) = S_{ij}(f_j) = \mathfrak{F}(S_{ij})(F_j) = U_{ij}(F_j) = V^{(k)}(F).$$

Thus, $\boldsymbol{F}^N(T^{(k)}) = V^{(k)}$. By the linearity of $\boldsymbol{F}^N$ we get $\boldsymbol{F}^N(S) = (\mathfrak{F}(\sum V^{(k)}))$.

By a similar reasoning, if $\mathfrak{P} = (P_{ij})$ where the P_{ij} are polynomials of derivation, then for $f \in D^N$, $\boldsymbol{F}^N(\mathfrak{P}f) = \mathfrak{Q}F$, where $\mathfrak{Q}$ is the matrix of polynomials Q_{ij} such that, for $g \in \mathfrak{D}$, $\mathfrak{F}(P_{ij}g) = Q_{ij}G$, and where, for $G \in \mathfrak{D}^N$,

$$\mathfrak{Q}G = (\sum_{j=1}^{N} Q_{ij}G_j, \cdots, \sum_{j=1}^{N} Q_{Nj}G_j).$$

For any $z \in C$, we shall denote by $\mathfrak{Q}(z)$ the matrix $(Q_{ij}(z))$. We shall write, for convenience, $\mathfrak{Q} = \boldsymbol{F}^N(\mathfrak{P})$ when $\mathfrak{Q}$ and $\mathfrak{P}$ are so related. Conversely, given any matrix $\mathfrak{Q}$ of polynomials, we can find a matrix $\mathfrak{P}$ of polynomials of derivation such that $\mathfrak{Q} = \boldsymbol{F}(\mathfrak{P})$.

If $\mathfrak{P} = (P_{ij})$ where the P_{ij} are constant-coefficient partial differential operators on $\mathfrak{D}$, we can form the determinant of $\mathfrak{P}$, which is a constant coefficient partial differential operator on $\mathfrak{D}$. We say that $\mathfrak{P}$ is *degenerate* if this operator is 0; otherwise we call $\mathfrak{P}$ *non-degenerate*. If $\mathfrak{Q}$ is a matrix of polynomials in n-letters, then we can form the determinant of $\mathfrak{Q}$ which is a polynomial. We call $\mathfrak{Q}$ *degenerate* if this polynomial is zero; if it is different

from zero, we say $\mathfrak{Q}$ is *non-degenerate*. From the above it is clear that, for $\mathfrak{P}$ a matrix of polynomials of derivation on $\mathcal{D}$, and $\mathfrak{Q} = \boldsymbol{F}^N(\mathfrak{P})$, then $\mathfrak{Q}$ is degenerate if and only if $\mathfrak{P}$ is and, even $\boldsymbol{F}$ (determinant $\mathfrak{P}$) = determinant $\mathfrak{Q}$.

For any $F \in \mathcal{D}^N$, $z \in C$, we define

$$\| F(z) \| = (\| F \|)(z) = (|F_1(z)|^2 + |F_2(z)|^2 + \cdots + F_N(z)|^2)^{\frac{1}{2}}.$$

Theorem 1 has the consequence:

THEOREM 10. *Let $l > 0$ and $\{F_i\}$ be a sequence in $\mathcal{D}^N$. A necessary and sufficient condition that $F_i \to 0$ in $\mathcal{D}^N$ is: For any $r > 0$ and any polynomial Q on C, $|Q(z)| \, \| F_i(z) \| \to 0$, uniformly for $z \in C_r$.*

It is possible to carry out the procedure of Section 4 and obtain the analogue for $\mathcal{D}^N$ of Theorem 5. In order to do this, we had originally used a method of estimating the size of the characteristic roots of a matrix of polynomials in n-variables. However, following the suggestion of the referee, we shall adopt an algebraic approach.

Let $\mathfrak{P} = (P_{ij})$ where the P_{ij} are polynomials of derivation in n variables; assume that P is non-degenerate. Then we have, for any $f \in \mathcal{D}^N$, if $\mathfrak{P}f = (g_1, g_2, \cdots, g_N)$, $\sum_{j=1}^{N} P_{ij} f_j = g_i$, for $i = 1, 2, \cdots, N$. Thus, by Cramer's rule, if Δ denotes the determinant of $\mathfrak{P}$, and Δ_{ij} the cofactor of P_{ij},

$$\Delta f_j = \sum_{i=1}^{N} \Delta_{ij} g_i, \tag{7.1}$$

for $j = 1, 2, \cdots, N$.

Let us denote by $\mathfrak{P}\mathcal{D}^N$ the space of $\mathfrak{P}h$ for $h \in \mathcal{D}^N$; the topology of $\mathfrak{P}\mathcal{D}^N$ is that induced by $\mathcal{D}^N$. The spaces $\mathfrak{P}\mathcal{D}_l^N$ for $l > 0$ are defined similarly. Then (7.1) together with the corollary to Theorem 5 give immediately

THEOREM 11. *Let $\mathfrak{P}$ be a non-degenerate matrix of polynomials of derivation in n variables. Then $\mathfrak{P}f \to f$ is a continuous linear map of $P\mathcal{D}_l^N \to \mathcal{D}_l^N$ for each $l > 0$.*

Let us denote by $\mathcal{D}_F^N$ the space of elements of $\mathcal{D}^N$ with the topology of coordinatewise convergence in $\mathcal{D}_F$. By $\mathcal{D}_F^{*N}$ we denote the set of mappings of $\mathcal{D}_F^N$ into the complex N-space. Each $S \in \mathcal{D}_F^{*N}$ can be represented by a matrix (S_{ij}) where each $S_{ij} \in \mathcal{D}_F'$.

THEOREM 12. *Any non-degenerate matrix $\mathfrak{P}$ of polynomials of derivation is completely inversible in $\mathcal{D}_F^{*N}$, that is, for any $S \in \mathcal{D}_F^{*N}$ there is a $T \in \mathcal{D}_F^{*N}$ such that $\mathfrak{P}'T = S$. In particular, $\mathfrak{P}$ has an elementary solution.*

Proof. This can be proven in a manner similar to the proof of Theorem 6, but we shall use a different method because we can use it for further results. Let $U = (U_{ij}) \varepsilon \mathcal{D}_F{}^{*N}$ where each $U_{ij} \varepsilon \mathcal{D}_F'$. Then we have $(P'U)_{ij} = \sum_{k=1}^{N} P_{kj}' U_{ik}$ because, if $h = (h_2, \cdots, h_N) \varepsilon \mathcal{D}^N$,

$$P'U \cdot h = U \cdot Ph = \sum_{k=1}^{N} U_{ik} \cdot \sum_{j=1}^{N} P_{kj} h_j = \sum_{j=1}^{N} \sum_{k=1}^{N} P'_{kj} U_{ik} \cdot h_j,$$

which is the stated result.

Now, if we call $U_b = U_{ij}$ where $b = i + N(j-1)$, then it is easily seen that the transformation $\mathfrak{P}'$ on the space of (U_b) is of the form

$$\boldsymbol{P} = \begin{pmatrix} \mathfrak{P}' & & & \\ & \mathfrak{P}' & & 0 \\ & & \ddots & \\ & 0 & & \mathfrak{P}' \end{pmatrix},$$

where $\mathfrak{P}'$ is the matrix (P_{ij}'). Since, for D_1, D_2 any polynomials of derivation, $(D_1 D_2)' = D_1' D_2'$, the determinant D of $\boldsymbol{P}$ is different from zero. Let us define V_b for $b = 1, 2, \cdots, N^2$ by

$$V_b = \sum_{a=1}^{N^2} D_{ab} S_a \tag{7.2}$$

where D_{ab} is the cofactor of the term in the a row, b column of $\boldsymbol{P}$. By Theorem 6 we can find $T_b \varepsilon \mathcal{D}'$ such that $D' T_b = V_b$. By Cramer's rule, $\mathfrak{P}' T = S$, which is the desired result.

By $\mathfrak{P}\mathcal{D}_F{}^N$ we denote the space of $\mathfrak{P}f$ for $f \varepsilon \mathcal{D}_F{}^N$ with the topology induced by $\mathcal{D}_F{}^N$. Exactly as in the proof of Theorem 11, we can show

THEOREM 13. *With $\mathfrak{P}$ as above, $\mathfrak{P}f \to f$ is a continuous linear map of $\mathfrak{P}\mathcal{D}_F{}^N \to \mathcal{D}_F{}^N$.*

The spaces $\mathcal{E}^N$, $\mathcal{E}_F{}^{*N}$ are defined similarly. By $\mathfrak{P}'\mathcal{E}_F{}^{*N}$ we denote the space of all $\mathfrak{P}'S$ with $S \varepsilon \mathcal{E}_F{}^{*N}$, the topology being that induced by $\mathcal{E}_F{}^{*N}$.[3] Then we have, as in the proofs of Theorem 11, 12:

THEOREM 14. *$\mathfrak{P}'S \to S$ is a continuous linear map of $\mathfrak{P}'\mathcal{E}_F{}^{*N} \to \mathcal{E}_F{}^{*N}$.*

THEOREM 15. *For any $f \varepsilon \mathcal{E}^N$ there is a $g \varepsilon \mathcal{E}^N$ which satisfies $\mathfrak{P}g = f$.*

[3] Footnote added in proof: Again it can be shown (see footnote 2) that $\mathcal{E}_F{}^{*N}$ is the same as $\mathcal{E}^{*N}$ (which is defined in the obvious manner).

Suppose now that $\mathfrak{P}$ is degenerate. Since the ring of polynomials of derivation has no zero divisors, we can procede along usual lines (see [7]) to obtain an $f \in \mathcal{D}^N$ such that $\mathfrak{P}f = 0$. Moreover, we can find a polynomial of derivation Y such that every Yf for $f \in \mathcal{D}^N$ can be written uniquely in the form $f' + f''$ where $f' \in \mathcal{N}'$ (null space of $\mathfrak{P}$) and $f'' \in \mathcal{N}''$ (on which $\mathfrak{P}$ acts in a non-degenrate manner). Let $\mathfrak{P}\mathcal{D}_F{}^N$ be the space of $\mathfrak{P}f$, $f \in \mathcal{D}^N$, with the topology induced by $\mathcal{D}_F{}^N$. Then we get, in the usual manner,

THEOREM 16. *$\mathfrak{P}f \to f''$ is a continuous linear map of $\mathfrak{P}\mathcal{D}_F{}^N \to \mathcal{D}_F{}^N$.*

We can also prove, by conventional methods

THEOREM 17. *Let $S \in \mathcal{D}_F{}^{*N}$; a necessary and sufficient condition that there exist a $T \in \mathcal{D}_F{}^{*N}$ satisfying $\mathfrak{P}'T = S$ is that $S = 0$ on $\mathcal{N}'$.*

COROLLARY. *If $\mathfrak{P}$ is degenerate, there is no elementary solution for $\mathfrak{P}$.*

Proof. Let $f \in \mathcal{D}^N$ satisfy $\mathfrak{P}f = 0$. Since derivation and translation commute, we may assume $f(0) \neq 0$. Thus,

$$\delta^N \cdot f = (f_1(0), f_2(0), \cdots, f_N(0)) \neq 0,$$

and the result follows from Theorem 17.

We can prove similar results for $\mathcal{E}^N$.

8. General remarks. In Section 4 we proved that $L: P\mathcal{D}_F \to \mathcal{D}_F$ defined by $L(Pf) = f$ for all $f \in \mathcal{D}_F$ is a continuous linear map. In case that $n = 1$ and P is an ordinary differential operator with constant coefficients, then Schwartz [5] has shown that L can be extended to a continuous linear map $M: \mathcal{D}_F \to \mathcal{D}_F$. This results from the fact that the quotient space $\mathcal{D}_F/P\mathcal{D}_F$ is finite dimensional. Such is no longer the case for $n > 1$ and P arbitrary. We do not know whether L can be extended in the general case. Such an extension would be a generalized integral.

The above question can be reformulated as follows: Call $(P\mathcal{D}_F)'$ the dual of $P\mathcal{D}$ with the topology of uniform convergence on the bounded sets of $P\mathcal{D}_F$. Then the adjoint L' of L is a continuous linear map of $\mathcal{D}_F' \to (P\mathcal{D})_F'$. ($L'$ is defined by $(L'S)(Pg) = S(LPg)$ for any $g \in \mathcal{D}_F$.) Can we *lift* L' to a continuous linear map of $\mathcal{D}_F' \to \mathcal{D}_F'$? Or, what is the same thing: We know that every element of $(P\mathcal{D})_F'$ can be extended to an element of $\mathcal{D}_F'$; does there exist a canonical way of extending the element of $(P\mathcal{D})_F'$ to $\mathcal{D}_F'$? That is, can we find a continuous linear map $M: (P\mathcal{D})_F' \to \mathcal{D}_F'$ such that, for $T \in (P\mathcal{D})_F'$, $M(T)$ is an extension of T? We hope to discuss this question at a future date.

Similar remarks apply to the space $\mathcal{E}_F$ (Theorem 9).

An interesting question is raised in the statement of Lemma 3. Let us denote, more precisely, by $\alpha(m)$ the constant α which arises in that Lemma. What is the best (greatest) possible value of $\alpha(m)$ as a function of m? The question may be reformulated as follows: Given m points $x_1, \cdots, x_m$ in the unit interval, what is the largest possible value of $\alpha(m)$ for which we can be sure of finding an x in the unit interval such that $\sum_{i=1}^{m} | x - x_i | \geqq \alpha(m)$. The methods of our proof of Lemma 3 which, of course, were not designed to find a best possible result, show that we may choose $\alpha(m) = 1/2^m(m+1)^m$. My friend, Dr. Phlotz Newman, has pointed out to me that the best possible result is $\alpha(m) = 2/4^m$ (see Polyá and Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925.)

All the theorems on division that we have proven involve the division by a polynomial in a space of analytic functions. An important question is the following: If P is a polynomial in n-variables and $S \in \mathcal{D}'$, can we find a $T \in \mathcal{D}'$ such that $PT = S$? It is clear that our methods shed no light on this problem.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] S. Banach, *Théorie des Opérations Linéares*, Warsaw, 1932.

[2] J. Dieudonné and L. Schwartz, "La Dualité dans les Espaces ($\mathcal{F}$) et ($\mathcal{LF}$)," *Ann. Inst. Fourier Grenoble*, vol. 1 (1950), pp. 61-101.

[3a] B. Malgrange, "Équations aux dérivées partielles à coefficients constants. 1. Solution élémentaire," *Comptes rendus Academie Sciences*, vol. 237 (1953), pp. 1620-1622.

[3b] ———, "2. Équations avec second membre," *ibid.*, vol. 238 (1954), pp. 196-198.

[4] R. E. A. C. Paley and N. Wiener, *Fourier Transforms on the Complex Domain*, New York, 1934.

[5] L. Schwartz, *Théorie des Distributions*, vol. I and II, Paris, 1951.

[6] E. C. Titchmarsh, *The Theory of Functions*, Oxford, 1932.

[7] B. L. Van der Waerden, *Moderne Algebra*, vol. II, Berlin, 1931.

A PERIODIC ALGORITHM FOR CUBIC FORMS II.*

By HARVEY COHN.

15. Introduction to Part II. In this continuation of our discussion of a cubic sign-discrimination algorithm, introduced in an earlier volume of this JOURNAL [7], we shall turn our attention to the computation of units.

First of all we shall show (leaning once more on the Minkowskian algorithm) that all units are discoverable by the "neighbor process" of Part I [7], pp. 829-833. In so doing we must restrict our lattice to a module $\mathfrak{M}$ of preferably (but not necessarily) all integers in a totally real cubic field $\mathfrak{K}$. The field $\mathfrak{K}$ need not be abelian but the module $\mathfrak{M}$ must be non-singular (i. e., of dimension three over the rationals). The units would then lie in $\mathfrak{O}$ the order of the module $\mathfrak{M}$.

Secondly we shall consider the manner in which the units should be written. As we observed in [7], p. 833 certain 3×4 matrices are proportional only under certain rearrangements of rows and columns. This makes it desirable to have a hypercomplex notation, which also brings out normal symmetries of the matrices.

In comparing our modified algorithm with its Minkowskian progenitor we start with the realization that what we are doing is refining the algebraic properties of the algorithm at the expense of the geometric properties, but the algebraic simplifications make numerical work [1] more feasible while the geometric complications make themselves felt only in existence type proofs which are disposed of here. This dichotomy is illustrated in the geometric and algebraic portions of the present work.

Existence of Units.

16. Types of neighbors. We consider a reduced matrix ϕ in the sense of **4** (above) consisting of four column vectors which we shall write as

* Received February 5, 1954. Research sponsored by the U. S. Army Office of Ordnance Research under Contract No. DA-20-018-ORD-12332. The sections and bibliographical refernces are numbered consecutively with those in item [7] which formed "Part I" of the present work. Errata for Part I are appended to the present paper.

[1] The present method is being programmed for the EDVAC at Aberdeen, Md., by Dr. Saul Gorn.

$$\phi = [P, Q, R, S]. \tag{16.1}$$

We shall classify the 24 neighbor possibilities (explained in **10** above) into three types depending on the *partition* of the symbols serving as *second* and *third* element in the neighbor operation:

(16.2)

Type 1	Type 2	Type 3
P, S or Q, R	Q, S or P, R	R, S or P, Q
$(QPS), (PQR)$	$(PQS), (QPR)$	$(PRS), (RPQ)$
$(RPS), (SQR)$	$(RQS), (SPR)$	$(QRS), (SPQ)$
$(QSP), (PRQ)$	$(PSQ), (QRP)$	$(PSR), (RQP)$
$(RSP), (SRQ)$	$(RSQ), (SRP)$	$(QSR), (SQP)$

LEMMA. *Every reduced matrix* ϕ *has either one or two* (*reduced*) *neighbors of each type.*

Outline of Proof. We must consider the subdivision into cases as determined by the relative magnitudes of each row: $(|p_i|, |q_i|, |r_i|, |s_i|)$, (compare **11**). There are actually twenty-four different cases (to within elementary interchanges of rows and columns and changes of sign in rows). The details are omitted in the interests of brevity we simply remark that this method was used in the (weaker) statement made earlier [7], p. 830 that there are from three to six neighbors.

17. Occurrence of independent units. The units deduced from the Minkowskian algorithm were "controllable" in the sense that if we take a succession of neighbors to one Minkowskian triple by "lowering the same face" (say x_i) each time we would obtain a unit whose x_i conjugate is less than unity and whose other conjugates are greater than unity in absolute value. If units are formed corresponding to two different x_i they are independent by the classical argument of Dirichlet.

The same thing can be done here, except that the "face" corresponding to x_i belongs to none of the vectors in the quadruple (16.1), e. g., it can happen that the maximum of $|p_1|, |q_1|, |r_1|, |s_1|$ cannot be decreased through any neighbor operation. We need to use other quantities namely

$$h_i = u_i^{(i)}, \qquad 1 \leq i \leq 3, \tag{17.1}$$

where $u_i^{(j)}$ are the components of $U^{(j)}$:

$$U^{(1)} = P + S, \qquad U^{(2)} = Q + S, \qquad U^{(3)} = R + S. \tag{17.2}$$

(The h_i are positive if we suppress the "$\pm$" in the left hand side of (4.22) as we now do for convenience).

LEMMA. *A neighbor operation of type i decreases h_i while it increases (or leaves the same) the other two h_j $(j \neq i)$.*

Proof. It suffices by the very abundant symmetry to note that the operation (QPS) of type 1 changes h_1 $(= p_1 + s_1)$ into s_1 $(< h_1)$ while it changes h_3 $(= r_3 + s_3)$ into $r_3 + s_3 - p_3$ $(> h_3)$ and leaves h_2 $(= q_2 + s_3)$ unchanged. Q. E. D.

We note that the triple $(U^{(1)}, U^{(2)}, U^{(3)})$ is not Minkowskian since it determines a sub-lattice of index 2 but it is a kind of "invariant" for the quadruple (16. 1) by equation (4. 21).

Thus by virtue of the finiteness theorem of **8** (above) units are determined which are independent. (Compare [3], p. 444).

18. Occurrence of all units. Consider next how an *arbitrary* unit in $\mathfrak{O}$, namely $E = (e_1, e_2, e_3)$, (taken with positive norm for convenience), can be discovered by our algorithm. If we form the product

$$\phi E = [PE, QE, RE, SE] \tag{18. 1}$$

we obtain a quadruple of vectors lying in $\mathfrak{M}$ (by definition of $\mathfrak{O}$), and spanning $\mathfrak{M}$ again (since the sublattice spanned by ϕE has determinant equal to that of ϕ times the norm of E). By a rearrangement of columns (vectors) in ϕE we therefore obtain another reduced quadruple. Therefore all we need show is that ϕ and ϕE are connected by a chain of neighbor operations (as in **12** above).

We introduce the term *overlapping* to describe two sets of vectors such as triples or quadruples, etc., with *two* or more common elements.

LEMMA 18. 1. *Two overlapping quadruples can be connected by means of neighbor operations.*

Outline of proof. Let the two overlapping quadruples be $\phi_{0,0}$ and $\phi_{a,b}$ (for integral a, b) where

$$\begin{aligned} &\phi_{a,b} = [P, Q, R^{(a,b)}, S^{(a,b)}] \\ &R^{(a,b)} = R + aP + bQ, \qquad S^{(a,b)} = S - aP - bQ. \end{aligned} \tag{18. 2}$$

Now the requirement that $\phi_{a,b}$ be reduced leads to three sign conditions on $R^{(a,b)}$ and $S^{(a,b)}$ $(= -R - (a+1)P - (b+1)Q)$. Thus

$$\begin{aligned} f_1(a+1, b+1) < 0 \qquad & f_1(a, b) < 0 \\ f_2(a+1, b+1) < 0 \qquad & f_2(a, b) < 0 \\ f_3(a+1, b+1) < 0 \qquad & f_3(a, b) > 0, \end{aligned} \tag{18. 3}$$

where (under restrictions (4. 21) and (4. 22)),

$$f_i(a, b) = r_i + ap_i + bq_i. \tag{18. 4}$$

It is then a tedious but elementary matter to show (by representing the six inequalities (18. 3) graphically) that if the relations (18. 3) hold for (a, b) then a and b are not both of the same sign, while the inequalities (18. 3) hold with (a, b) replaced by $(a, b + \operatorname{sgn} a)$ or $(a + \operatorname{sgn} b, b)$. (Here $\operatorname{sgn} 0 = 0$). Thus Lemma 18. 1 follows by descent. Q. E. D.

LEMMA 18. 2. *To any reduced quadruple (or to one connected with it by a chain of neighbors) there corresponds at least one overlapping Minkowskian basis.*

The proof of this lemma will be deferred to the next section; (see Lemma 19. 3 below).

LEMMA 18. 3. *Any two neighboring Minkowskian bases each overlap overlapping reduced quadruples.*

For it is the nature of Minkowski's neighbor concept, [3], that any two neighboring triples overlap; but by the theorem in **5** these triples are contained (discounting sign) in a reduced quadruple if they are bases.

Now we use the further information that any two Minkowskian triples are connected by the Minkowski neighbor process. The proof is based on a topological argument suggested by Minkowski (and proved in [2], p. 342). Since every degenerate Minkowskian triple has non-degenerate triples (i. e., basis triples) as neighbors, it takes very little effort to deduce from the same proof that any two Minkowskian bases are connected by a chain of Minkowskian bases as neighbors. Hence the lemmas yield the following result and with it the "discoverability" of all units:

THEOREM. *Any two reduced quadruples can be connected by a chain of neighboring (reduced) quadruples.*

19. Proximity of algorithms. In Section **5** (above) we showed that every Minkowskian basis is part of a reduced quadruple, which means, in effect, that every *Minkowskian triple overlaps a reduced quadruple.* The "converse" seems to be false, indicating that there are "more" reduced quadruples than Minkowskian triples in some sense perhaps analogous to the sense that the quadratic algorithm we introduced in **2** (above), possessing "mediants" in addition to regular convergents of the simple continued fraction contains "more" reduced triples (see Section **25** below). Lemma 18. 2

seems to be as close as we can comfortably approach a converse to the theorem in **5**.

To get at the difficulty we define *a mediant quadruple to be a reduced quadruple in which one vector achieves in each of its three components the minimum absolute value for all four vectors of the quadruple.* Thus if (say)

$$|r_i| < \min |p_i|, |q_i|, |s_i|, \qquad 1 \leq i \leq 3, \tag{19.1}$$

then $[P, Q, R, S]$ is a mediant quadruple. Here we would call R the *small vector.*

We call a set of vectors $\mathfrak{S}$ *limiting* if each (non-zero) vector achieves an absolute value in some x_i-component not exceeding that of a vector of $\mathfrak{S}$. Then by the Minkowskian method of admissible bodies if a set of limiting vectors contains at least three vectors none of which can be omitted from $\mathfrak{S}$ (without loss of limiting property) then a Minkowskian triple can be chosen from the set $\mathfrak{S}$. If $\mathfrak{S}$ consists of two vectors none of which can be omitted from $\mathfrak{S}$ (without loss) then a Minkowskian triple can be chosen that overlaps $\mathfrak{S}$.

Now mediant quadruples are undesirable because they do not have two vectors that serve as a limiting set (by virtue of the fact that the *small vector* can be omitted without loss of limiting property). Hence we must use

LEMMA 19.1. *No complete chain of reduced neighboring quadruples can consist entirely of mediants.*

Proof. Suppose R is the small vector of a mediant ϕ. Then any neighbor operation with R "in the middle" is valid. Take in particular (PRS): We can apply the neighbor operation only a limited number of times, e. g.,

$$\phi(PRS)\phi'(PRS)\phi'' \cdots \phi^{(n)}, \tag{19.2}$$

(since the operation decreases p_1 (>0) by $|r_1|$ each time). Now if $\phi^{(n)}$ is still a mediant, the small vector must be Q (because it is no longer "R" and yet $|p_2|$ and $|s_2|$ now exceed $|r_2|$). We therefore apply (PQS) as long as possible until we obtain a $\phi^{(nm)}$. This process must terminate since the operations (PQS) and (PRS) each decrease p_1 (>0) by $|q_1|$ or $|r_1|$ each time. Q. E. D.

We are now ready to handle the non-mediant quadruples. We are again confronted with a multitude of special cases that would be tedious to give in detail. We therefore first make use of limiting sets.

LEMMA 19.2. *For any reduced quadruple the twenty-six vectors formed*

from the following and their negatives are a limiting set:

$$P, Q, R, S, \quad P-Q, \quad P-R, \quad P-S, \quad Q-R, \quad Q-S, \quad R-S, \quad P+S, \quad Q+S, \quad R+S. \tag{19.3}$$

Proof. Let $V = aP + bQ + cR$. Since $V = (b-a)Q + (c-a)R - aS$ by (4.21), we can assume a, b, c are not all of the same sign. We therefore can select a component j for which $|v_j| = |ap_j| + |bq_j| + |cr_j|$. Hence the result. Q. E. D.

LEMMA 19.3. *If a reduced quadruple is not a mediant then it contains a limiting set of two vectors (and therefore overlaps a Minkowskian basis).*

Outline of proof. An actual listing of the twenty-four different types of ϕ according to the relative magnitudes of the rows (see **16**), would show that only two types are mediants. The remaining twenty-two types reveal limiting sets on inspection, but the amount of space required to handle the inequalities in writing is almost prohibitive. It should be clear however that Lemma 19.2 expresses the problem in finite terms.

Hypercomplex Notation and Normal Symmetries.

20. Reduction redefined. The distinguishing algebraic features of our present algorithm can best be brought out by considering the hypercomplex system $\mathfrak{A}$ over the module $\mathfrak{M}$ generated by the introduction of elements $(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, 1)$ that form the four-group:

$$\boldsymbol{a}_i^2 = \mathbf{1}(= \boldsymbol{a}_0), \quad \boldsymbol{a}_i\boldsymbol{a}_j = \boldsymbol{a}_j\boldsymbol{a}_i = \boldsymbol{a}_k, \qquad 1 \leq i \neq j \neq k \leq 3. \tag{20.1}$$

Then each 3×4 matrix ϕ can be written in the form

$$\phi = P\boldsymbol{a}_1 + Q\boldsymbol{a}_2 + R\boldsymbol{a}_3 + S\mathbf{1}. \tag{20.2}$$

In the system $\mathfrak{A}$ the "scalars" are elements of $\mathfrak{M}$ for which each conjugate multiplies (row by row) the 3×4 matrix ϕ. Thus if $\mathfrak{M}$ were an order, $\mathfrak{A}$ would be an "algebra."

Now the element ϕ of $\mathfrak{A}$ will be called a reduced basis of $\mathfrak{M}$ (see 4 above) if any three of the quantities P, Q, R, S form a basis of $\mathfrak{M}$ and (see (4.21)) if ϕ goes to 0 by the mapping of $\mathfrak{A}$ that brings each $\boldsymbol{a}_i$ into $\mathbf{1}$ and if (continuing to suppress the "$\pm$" on the left of (4.22)), each x_i conjugate of P, Q, R, S agrees in sign with χ_i of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \mathbf{1}$ respectively, where the χ_i $(i = 1, 2, 3)$ are the three-non principal characters of the four-group, fixed so that $\chi_i(\boldsymbol{a}_i) = 1$.

Thus the finiteness theorem of **8** (above) asserts that *the system may have only a finite number of reduced basis elements to within (scalar) proportionality constants.* The argument in **9** enables us to choose these constants as units.

The hypercomplex notation fits in with the classification into neighbor types. If we use the symbol $(\mathbf{1}a_3a_2)$ to denote, for example, (SRQ), then a neighbor of type i (here type 1), is one in which the product of the second and third elements of the neighbor operation (e. g., $a_3 \cdot a_2$), equals a_i (here a_1). The new notation enables us to obtain another very curious kind of result that we mention in passing: The sum or product of any two reduced elements ϕ_1, ϕ_2 belonging to two modules in $\mathfrak{R}$ is equal to reduced elements belonging to a third module (which depends on the operation and elements).

Our main purpose in using the notation will be for dealing with units. (Compare the hypercomplex notation use by Hasse [8] for a somewhat different analysis of cubic units).

21. Hypercomplex units. Now we shall designate the units $E=(e_1, e_2, e_3)$ of $\mathfrak{O}$ that are not totally positive as multipliers in $\mathfrak{A}$. What we do is select the character χ_j for which $\chi_j(a_i)$, $(1 \leq i \leq 3)$, agrees in sign with e_i (assuming the norm of E positive as before). Then the product a_jE denotes the hypercomplex unit corresponding to E and for any reduced basis element ϕ of $\mathfrak{M}$, the product ϕa_nE yields another reduced basis element. This process would reduce to ordinary (scalar) multiplication for E totally positive or totally negative.

Now let $\mathfrak{G}$ be the (abelian) group of all hypercomplex units so defined. Let $\mathfrak{G}_0$ be the sub-group of totally positive units. Then the nature of the quotient group $\mathfrak{G}/\mathfrak{G}_0$ is uniquely determined by r the *signature rank,* i. e., the number of different arrays of signs that can be achieved by the scalar units (of positive and negative norms). Then for

$$(21.1) \qquad r=\begin{cases}8,\\4,\\2,\end{cases} \quad \mathfrak{G}/\mathfrak{G}_0=\begin{cases}[a_1E^{(1)}, a_2E^{(2)}, a_3E^{(3)}, 1],\\[a_iE^{(n)}, 1],\\[1].\end{cases}$$

Note that by the eigenvalue discussion in **14** (above), only these hypercomplex *units* can emerge from proportionalities of reduced basis elements under interchange of columns with rows preserved.

22. Mixing of rows. Consider two reduced basis elements ϕ and ϕ^* of $\mathfrak{A}$. We are now concerned with proportionalities of one row of ϕ with another row of ϕ^*, accompanied by the necessary permutation of columns.

For instance, a (deceivingly trivial) illustration is provided by $\phi = \phi^* = \phi_0$ of **11** (above), where the first row is proportional to the second (with ratio 1) under cyclic permutation of P, Q, R. Such proportionality evidently could occur only for $\mathfrak{K}$ abelian since otherwise p_1/q_1 could equal (say) q_2^*/r_2^* only when both ratios (belonging to conjugate fields) are rational, in contradiction to the basis property. Actually we can say more:

THEOREM. *Two reduced basis elements of $\mathfrak{M}$ can have different rows of each proportional (with some permutation of columns) if and only if the module $\mathfrak{M}$ is equivalent to a normal module.*

Proof. Suppose such a proportionality of rows exists. Then the second row of ϕ^* is (say) k_1 times the first row of ϕ (with rearrangement of columns), where k_1 is an element of the (abelian) field $\mathfrak{K}$. Likewise the third and first rows of ϕ^* (by using conjugates) are seen to be k_2 and k_3 times the second and third rows of ϕ (with rearrangement of columns). Thus $K = (k_1, k_2, k_3)$ is an element of $\mathfrak{K}$ with norm $N(K) = 1$ by virtue of the fact that ϕ and ϕ^* are basis elements of the same module. But by a well-known theorem of Hilbert [9], k_1, having norm 1, will be the ratio of conjugate elements, which (by multiplication with a rational integer) can be chosen to lie in $\mathfrak{O}$. Thus $k_1 = t_1/t_2$ where $T = (t_1, t_2, t_3)$ is in $\mathfrak{O}$ and has positive norm. Hence $\phi \boldsymbol{a}_i T$ and $\phi^* \boldsymbol{a}_i T$ are for some i $(0 \leq i \leq 3)$, reduced basis elements of system $\mathfrak{A}T$ corresponding to $\mathfrak{M}T = \mathfrak{N}$. Now $\mathfrak{N}$ is easily the normal module required. The converse is immediate. Q. E. D.

COROLLARY. *A single reduced basis element ϕ of $\mathfrak{M}$ will have one row proportional to another row (with permutation of columns) if and only if the module $\mathfrak{M}$ is equivalent to a module $\mathfrak{N}$ with a normal basis that furthermore consists of the conjugates of an element P of $\mathfrak{N}$ with positive norm and negative trace.*

The preceding proof applies except that we display the basis element by letting Q and R be conjugates of P and letting S be the negative of the trace of P.

The theorem and corollary are (strictly) more restrictive, i. e., not all abelian modules are normal, nor are all normal modules equivalent to one with a normal basis, nor must the normal basis have the desired sign on its norm and trace.

As a final remark note that in both the theorem and corollary the order $\mathfrak{O}$ of $\mathfrak{M}$ is the same as that of $\mathfrak{N}$ and therefore a *normal* module in $\mathfrak{K}$. Hence by a well known Minkowskian technique (see [8]), it follows that funda-

mental units of $\mathfrak{O}$ can be taken as two conjugates of a single unit. Thus the signature rank r (of **21** above) may equal only 2 or 8 (but never 4).

23. Use of Galois group operator. We can now enlarge our (abelian) hypercomplex system to a non-abelian one in the following manner: Consider $\boldsymbol{Z}$ the Galois group operator for our *abelian* field K (so that $\boldsymbol{Z}^3 = \boldsymbol{1}$), Then

$$P\boldsymbol{Z} = (p_1, p_2, p_3)\boldsymbol{Z} = (p_2, p_3, p_1). \tag{23.1}$$

Now the neecssity of preserving the sign pattern requires that a permutation of rows be balanced by a permutation of columns as follows:

$$\phi\boldsymbol{Z} = (R\boldsymbol{Z})\boldsymbol{a}_1 + (P\boldsymbol{Z})\boldsymbol{a}_2 + (Q\boldsymbol{Z})\boldsymbol{a}_3 + (S\boldsymbol{Z})\boldsymbol{1}, \tag{23.2}$$

Formally, then, we *derive* the commutator rules

$$\boldsymbol{a}_1\boldsymbol{Z} = \boldsymbol{Z}\boldsymbol{a}_3, \quad \boldsymbol{a}_2\boldsymbol{Z} = \boldsymbol{Z}\boldsymbol{a}_3, \quad \boldsymbol{a}_3\boldsymbol{Z} = \boldsymbol{Z}\boldsymbol{a}_1, \quad \boldsymbol{1}\cdot\boldsymbol{Z} = \boldsymbol{Z}\cdot\boldsymbol{1}. \tag{23.3}$$

Thus the juxtaposition of symbols $\boldsymbol{Z}^t\boldsymbol{g}$ (where $0 \leq t \leq 2$ and $\boldsymbol{g}$ belongs to the four-group), will have meaning as a (right) operator on ϕ and will in fact be isomorphic to the alternating group on four elements.

Now if we consider a module equivalent to a normal module, then for every hypercomplex unit $\boldsymbol{a}_jE$ $(0 \leq j \leq 3)$, (of **21** above) we have *three* units belonging to the new hypercomplex system, namely $\boldsymbol{Z}^t\boldsymbol{a}_jE$ $(0 \leq t \leq 2)$. In this manner the group $\mathfrak{G}^*$ of hypercomplex units in the enlarged system is determined from the group $\mathfrak{G}_0$ of totally positive units as follows (compare (21.1)):

$$r = \begin{cases} 8 \\ 2, \end{cases} \mathfrak{G}^*/\mathfrak{G}_0 = \begin{cases} [\boldsymbol{1}, \boldsymbol{Z}, \boldsymbol{Z}^2] \times [\boldsymbol{a}_1E, \boldsymbol{a}_2E\boldsymbol{Z}, \boldsymbol{a}_3E\boldsymbol{Z}^2, \boldsymbol{1}] \\ [\boldsymbol{1}, \boldsymbol{Z}, \boldsymbol{Z}^2] \times [\boldsymbol{1}]. \end{cases} \tag{23.4}$$

Note that for $r = 8$ the fundamental units are shown as conjugates of a single unit while $r = 4$ does not occur, in accordance with the conclusion of **22** above. In this way we obtain a group $\mathfrak{G}^*$ whose multiplication rules can be given as follows:

$$(\boldsymbol{Z}^t\boldsymbol{a}_jE)(\boldsymbol{Z}^u\boldsymbol{a}_kE') = \boldsymbol{Z}^{t+u}\boldsymbol{a}_lE'' \tag{23.5}$$

where $E'' = (E\boldsymbol{Z}^u)E'$ and $\boldsymbol{a}_l = \boldsymbol{a}_k\boldsymbol{Z}^{-u}\boldsymbol{a}_j\boldsymbol{Z}^u$.

Now, the only possible proportionalities that mix rows are those coming from the group $\mathfrak{G}^*$, i. e., based on *units*, for the cube of any element of $\mathfrak{G}^*$ is an element of $\mathfrak{G}$, hence based on a scalar unit. (See section **21**).

24. Degenerate quaternary forms. On the basis of the last three sections we are in a position to discuss the quaternary form corresponding to ϕ namely

$$\Phi = \Pi(m_1 p_i + m_2 q_i + m_3 r_i + m_4 s_i), \qquad 1 \leq i \leq 3. \tag{24.1}$$

Several different Φ can correspond to a single ϕ under permutation of the (m_j), according to the signature rank r of the units in the order $\mathfrak{O}$.

Case I: $\mathfrak{M}$ *not equivalent to a normal module.* Here $r = 2, 4$, or 8 and every Φ is one of a set of $r/2$ (i. e., $r/2$ different "points" in the "chain" diagrams of **12**).

Case II: $\mathfrak{M}$ *equivalent to a normal module.* Here $r = 2$ or 8 and every Φ is one of a set of $3r/2$ *except* when the ϕ is proportional (row by row) to one in which three of the quantities P, Q, R, S are conjugates. (In the exceptional case Φ is one of a set of $r/2$ again).

In the table [7, p. 832] we find evidence of some of these cases for $r = 8$. Case I corresponds to $\Delta^2 = 148$, while Case II (or its exception) corresponds to the proportionality classes that occur by twelves (or by fours). Note that the module for $\Delta^2 = 196$ is neither an ideal nor normal, yet the correspondence "by fours" means it must be proportional to a normal module (with even a normal basis). The numerical work is being carried much further than in Part I, but it will be the subject of a separate report.

25. Concluding remarks. In Part I, Section **2**, a modified quadratic simple continued fraction algorithm was outlined. It turns out to be more conventional than first thought, in fact it is equivalent to the continued fraction algorithm with the "mediant" fractions thrown in. The geometrical construction given in **2** can be seen to be a variation of Smith's construction [10].

In Part III, the concluding portion, we shall treat the algorithm as a means of discovering "small norms" (analogously with the Pell Equation).

Errata for Part I (i. e., [7]).

Page 828 line 7, a should read a_1
line 18, $p_1 q_1$ should read $p_1 q_2$
830 line 16, -1.25 should read 1.25

832 line 4, $-\rho^2[=R]$ should read $-\rho$

833 lines 8-10, by twelves should read by at most twelves
by fours should read by at most fours.

WAYNE UNIVERSITY.

BIBLIOGRAPHY.

[6] G. Bullig, "Ein periodisches Verfahren zur Berechnung eines Systems von Grundeinheiten in den total reelen kubischen Körpern." *Abhandlungen aus dem Mathematischen Seminar der Hanseschen Universität*, vol. 12 (1938), p. 369.

[7] H. Cohn, "A periodic algorithm for cubic forms," this JOURNAL, vol. 74 (1952), pp. 821-833.

[8] H. Hasse, "Die Einheitsgruppe in einem total reelen nicht zyklischen kubischen Zahlkörper und in zugehörigen bikubischen Normalkörpern," *Archiv der Mathematik*, vol. 1, 1948/9, pp. 42-46.

[9] D. Hilbert, "Die Theorie der algebraischen Zahlkörper," *Jahresbericht der deutschen mathematiken Vereinigung*, vol. 4, 1894/5, p. 271.

[10] J. F. Koksma, "Diophantische Approximationen," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 4, Berlin, 1936, p. 38.

ERRATUM.

C. T. Rajagopal, "On Riesz summability and summability by Dirichlet's series: a further addendum and corrigendum," this JOURNAL, vol. 76, 252-258.

Page 256, line 9 from above insert, in front of "$\to 1$," the factor "$\exp -(r_0+1)$."